Algebra lineal

V. V. Voevodin



EDITORIAL MIR MOSCÚ


МИР

В. В. ВОЕВОДИН

ЛИНЕЙНАЯ АЛГЕБРА

ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА

V. V. Voevodin

Algebra lineal

Traducido al español
por el ingeniero
K. P. MEDKOV

Primera reimpresión

EDITORIAL MIR
MOSCÚ

Impreso en la URSS

Primera edición 1982

Primera reimpresión 1986

На испанском языке

INDICE

PRÓLOGO

El presente manual reúne en sí el curso ampliado del álgebra lineal y de la geometría analítica.

Unas cuantas palabras sobre este libro. Está destinado, principalmente, para aquellos que han elegido la matemática de cálculo como una disciplina esencial de su preparación. La enseñanza concerniente a dicha especialidad está ligada en muchos aspectos con los cursos tradicionales de las matemáticas. No obstante, los cambios, tanto en la metodología de exposición como en el contenido del curso, resultan imperativos.

Ya en las primeras conferencias los estudiantes empiezan a entrar en contacto con los elementos del álgebra lineal y de la geometría analítica. Estas primeras conferencias sirven también de origen de la formación de su concepción científica. Por esta razón, la asimilación futura por los estudiantes de toda la matemática de cálculo depende muchísimo del carácter y sucesión de la exposición del curso dado.

Existe toda una serie de buenos libros referentes al álgebra lineal y a la geometría analítica. Sin embargo, su empleo directo para los fines de enseñanza resulta algo embarazoso. La causa principal de ello radica, a nuestro parecer, en que los futuros especialistas en cálculos deben saber más datos del álgebra lineal que lo comunicado corrientemente en los libros de texto. Los estudiantes que se especializan en matemática de cálculo no sólo deben recibir una exposición estricta y sistemática de las bases del álgebra y geometría, sino que ya en el primer año deben tener contacto con la enorme riqueza de conocimientos que ha acumulado la matemática de cálculo.

El contacto con los problemas de cálculo permite acentuar con toda eficacia, en interés de la matemática de cálculo, todas las cuestiones de importancia que se estudian en el curso de conferencias, y establecer estrecha relación entre la teoría y los métodos numéricos del álgebra lineal. Como material de partida, para ello, sirven los hechos más sencillos de tales apartados como errores del redondeo, inestabilidad a las perturbaciones de muchos conceptos fundamentales del álgebra lineal, estabilidad de los sistemas ortonormalizados, espacios métricos y normalizados, descomposición singular, formas bilineales y su relación con los procesos de cálculo, etc.

Desde luego, la inclusión del material nuevo y suficientemente amplio es imposible sin una reconstrucción sustancial del curso tradicional. En el presente libro se hizo el intento de tal reconstrucción.

V. Voevodin

PARTE I ESPACIOS LINEALES

CAPÍTULO 1 CONJUNTOS, ELEMENTOS, OPERACIONES

§ 1. Conjuntos y elementos

En cualquier dominio de actividad siempre nos encontramos con la necesidad de considerar diversas totalidades de objetos unidos entre sí mediante un criterio común.

Así, por ejemplo, al estudiar la estructura de tal o cual mecanismo, podemos considerar la totalidad de todas las piezas del mismo. En este caso, como objeto suelto de la totalidad dada puede servir toda pieza del mecanismo, mientras que el criterio que une dichos objetos consiste en el hecho de que todos ellos pertenecen a un mecanismo bien determinado.

Hablando de la totalidad de puntos de una circunferencia en un plano, tenemos en cuenta, en esencia, los objetos —puntos del plano— unidos por la propiedad de que todos ellos son equidistantes respecto de cierto punto fijado.

Una totalidad de objetos unidos entre sí mediante un criterio común suele llamarse en las matemáticas *conjunto* y los propios objetos, *elementos* del conjunto. Al concepto de conjunto no se puede atribuir una definición rigurosa. Por supuesto, podemos decir (¡cómo lo hemos hecho!) que el conjunto es "una totalidad", "un sistema", "una clase", etc. Sin embargo, todo esto sería más bien el uso formal de la riqueza del vocabulario de la lengua.

Con el fin de determinar un concepto, es necesario, ante todo, indicar de qué modo dicho concepto está ligado con las nociones más generales. Resulta imposible hacerlo para el concepto de conjunto, puesto que en las matemáticas no existe para el conjunto un concepto más general. Por esta razón, en lugar de determinar el concepto de conjunto nos vemos obligados a ilustrarlo con unos ejemplos.

Uno de los métodos más sencillos de describir un conjunto radica en que se da la lista completa de los elementos que integran el conjunto. Por ejemplo, el conjunto de todos los libros de una biblioteca, accesibles al lector, está completamente determinado mediante las listas de los catálogos de biblioteca; el conjunto de todos los precios de mercancías está completamente determinado por la lista de precios, etc. No obstante, dicho método sólo es aplicable a los conjuntos *finitos*, es decir, a los conjuntos que contienen un número finito de elementos. Los conjuntos *infinitos*, es decir, aquellos que contienen un número infinito de elementos, no pueden ser determinados con la ayuda de una lista. ¿Cómo, por ejemplo, podríamos alistar todos los números reales?

En aquellos casos en que un conjunto no pueda ser prefijado con ayuda de una lista o no es cómodo hacerlo, éste se da por indicación de la propiedad *característica*, es decir, por indicación de tal propiedad que poseen los elementos del conjunto, y sólo ellos. Por ejemplo, en los problemas en que se determinan lugares geométricos la propiedad característica del conjunto de puntos, que sirve de solución para el propio problema, no es otra cosa que la totalidad de las condiciones que han de ser satisfechas por dichos puntos de acuerdo con los requisitos del problema.

La descripción de un conjunto puede ser muy sencilla y no despertar dificultades algunas. Por ejemplo, si hablamos de un conjunto que consiste en dos números, 1 y 2, está *claro* que ni el número 3 ni un cuaderno ni un automóvil pueden integrar dicho conjunto. En el caso general, sin embargo, la definición de los conjuntos por medio de sus propiedades características conduce, algunas veces, a ciertas complicaciones. Existen varias razones por las cuales surgen las complicaciones citadas.

Una de ellas puede consistir en que los conceptos empleados para la descripción de un conjunto no están suficientemente determinados. Convengamos, por ejemplo, en considerar el conjunto de todos los planetas en el sistema solar. ¿De qué se trata? Se conocen nueve planetas grandes. Pero alrededor del Sol giran también más de un mil de planetas pequeños, o los asteroides. Los diámetros de algunos de estos planetas miden centenares de kilómetros y hay planetas cuyos diámetros no superan 1 km. A medida que se perfeccionan los métodos de observación irán descubriéndose unos planetas cada vez más pequeños y, al fin y al cabo, surgirá la pregunta ¿cuáles de ellos son planetas pequeños y cuáles representan meteoritos y polvo cósmico?

No siempre las dificultades en determinar la composición de un conjunto dependen sólo de las razones indicadas. Ocurre a veces que los conjuntos que a primera vista parecen estar bien determinados resultan, de hecho, determinados muy mal o incluso no están determinados del todo. Supongamos, por ejemplo, que un conjunto consta

sólo de un número. Definamos este número como "un *número entero mínimo que no puede ser determinado con la ayuda de una frase que tenga menos de cien palabras rusas*". Consideraremos que se utilizan en el experimento sólo las palabras sacadas de cierto diccionario, como también sus formas gramaticales, y que, además, en el diccionario están contenidas palabras del tipo "uno", "dos", etc.

Hemos de notar que, por un lado, tal número no debe existir, pues se define por una frase, destacada más arriba con cursiva, que contiene menos de cien palabras, y por el sentido de esta frase el número no puede ser definido de la manera semejante. Pero, por otro lado, como el número de las palabras rusas utilizadas es limitado, quiere decir que hay números que no pueden ser definidos por una frase que tiene menos de cien palabras y, por consiguiente, entre dichos números existe uno que es mínimo.

En la rama de las matemáticas, llamada *teoría de conjuntos*, se han acumulado muchos ejemplos de que la definición de un conjunto entraña contradicciones interiores. El estudio del problema bajo qué condiciones esto puede tener lugar, ha conducido a las investigaciones profundas en el dominio de la lógica. No obstante, dejamos al lado estas investigaciones. En lo que sigue siempre supondremos que se consideran solamente los conjuntos que están precisamente definidos sin contradicciones algunas y cuya composición no despierta ninguna duda.

Como regla, los conjuntos se designarán con las letras latinas mayúsculas A, B, . . . , y sus elementos, con las letras minúsculas a, b, ... Escribiremos $x \in A$, si el elemento x pertenece al conjunto A, y $x \notin A$, si el elemento x no pertenece al conjunto A.

A veces será introducido en la consideración el así llamado conjunto *vacío*, es decir, un conjunto que no contiene ni un solo elemento. El uso del conjunto vacío resulta cómodo en el caso cuando se desconoce de antemano si existe aunque sea un sólo elemento en la totalidad que se considera.

Ejercicios.

1. Constrúyanse los ejemplos de conjuntos finitos e infinitos. ¿Qué propiedades son para ellos características?

2. Constrúyanse los ejemplos de conjuntos cuya descripción contiene una contradicción.

3. ¿Será vacío el conjunto de raíces reales del polinomio $x^4 + 4x^3 + 7x^2 + 4x + 1$?

4. Constrúyanse los ejemplos de conjuntos cuyos elementos son también conjuntos.

5. Constrúyanse los ejemplos de conjuntos que contienen a sí mismos a título de uno de sus elementos.

§ 2. Operación algebraica

Entre toda clase de conjuntos se pueden distinguir aquellos cuyos elementos admiten que se ejecuten con ellos ciertas operaciones. Supongamos, por ejemplo, que el objeto de nuestra consideración es un conjunto de todos los números reales. En este caso para todo elemento de dicho conjunto resultan definidas tales operaciones como el cálculo del módulo del elemento, el cálculo del seno del último; para cada par de elementos están definidas las operaciones de su adición y multiplicación.

En el ejemplo considerado prestaremos una atención especial a las siguientes peculiaridades de las operaciones mencionadas. En primer lugar, el carácter determinado de todas las operaciones para *cualesquiera* elementos del conjunto dado; en segundo, *la univocidad* de todas las operaciones y, por fin, la pertenencia del resultado, que se obtiene al realizar cualquiera de las operaciones, a los elementos *del mismo* conjunto. Tal situación no siempre tiene lugar.

Una operación puede ser definida no para todos los elementos del conjunto; por ejemplo, el cálculo de un logaritmo no está definido para los números negativos. La extracción de una raíz cuadrada de los números positivos está definida, mas de manera no unívoca. No obstante, incluso si una operación está unívocamente definida para todos los elementos, el resultado de su realización puede no constituir un elemento del conjunto dado. Consideraremos una operación de división en el conjunto de los números positivos enteros. Está claro que para cualesquiera dos números de este conjunto la operación de división es realizable, pero el resultado de su ejecución no será necesariamente un número entero.

Sea dado un conjunto A que contiene al menos un solo elemento. Diremos que en el conjunto A está definida *una operación algebraica*, si se indica una ley según la cual a todo par de elementos, a y b, tomados de dicho conjunto en el orden determinado, se le pone en correspondencia de manera unívoca el tercer elemento c que también pertenece al mismo conjunto.

Esta operación puede llamarse adición y en tal caso c será la suma de los elementos a y b, lo que se designará mediante el símbolo $c = a + b$; esta misma operación puede denominarse multiplicación y entonces c se llamará producto de los elementos a y b, lo que se designa con el símbolo $c = ab$.

En general, la terminología y la notación para la operación definida en el conjunto A no desempeñarán en adelante algún papel de importancia. Como regla, haremos uso de la notación de una suma o de un producto independientemente de cómo la operación está definida en realidad. Si, en cambio, nos hace falta subrayar ciertas propiedades generales de una operación algebraica, designaremos la operación con el símbolo $*$.

Recurriendo a unos ejemplos sencillos, veamos qué peculiaridades puede tener una operación algebraica. Supongamos que el conjunto A represente en sí una *totalidad* de todos los números racionales positivos. Introduzcamos para los elementos de este conjunto las operaciones ordinarias de multiplicación y *división* de los números y hagamos uso de la notación generalmente aceptada. No es difícil comprobar que ambas operaciones en el conjunto A son algebraicas. Sin embargo, si en la operación de multiplicación $ab = ba$ para todos los elementos de A, es decir, si la indicación del orden de los elementos no es esencial, para la operación de *división*, por el contrario, el orden de los elementos resulta muy esencial, pues la igualdad $a : b = b : a$ se verifica sólo en el caso en que $a = b$. De este modo, aunque la operación algebraica *está* definida para un par ordenado de elementos, la ordenación de los últimos no es, a veces, esencial.

Una operación algebraica se denomina *conmutativa*, si el resultado de su aplicación no depende del orden en que se eligen los elementos, es decir, si para cualesquiera dos elementos a y b del conjunto dado se verifica la igualdad $a \circ b = b \circ a$. Es evidente que entre las operaciones con los números universalmente admitidas la adición y la multiplicación son operaciones conmutativas, mientras que la sustracción y la división, operaciones no conmutativas.

Supongamos ahora que se consideran tres elementos arbitrarios a, b, c. Surge naturalmente la pregunta ¿qué sentido se atribuirá a la expresión $a \circ b \circ c$? ¿Cómo se debe aplicar una operación algebraica definida para dos elementos, a los tres elementos?

Puesto que una operación algebraica puede aplicarse solamente a un par de elementos, podemos atribuir un sentido determinado a la expresión $a \circ b \circ c$, encerrando entre paréntesis o bien dos primeros elementos, o bien dos elementos últimos. En el primer caso la expresión tomará la forma $(a \circ b) \circ c$; en el segundo caso, $a \circ (b \circ c)$. Examinemos los elementos $d = a \circ b$ y $e = b \circ c$. Debido a que ellos son elementos del conjunto inicial, las expresiones $(a \circ b) \circ c$ y $a \circ (b \circ c)$ pueden considerarse como resultado de aplicar una operación algebraica a los elementos d, c y a, e, *respectivamente*.

En general, los elementos $d \circ c$ y $a \circ e$ pueden resultar ser diferentes. Consideraremos de nuevo un conjunto de números racionales positivos y una operación algebraica de división de los números en este conjunto. Es fácil convencerse de que, como regla, se tiene $(a : b) : c \neq a : (b : c)$. Por ejemplo, $((3/2) : 3) : (3/4) = 2/3$, pero $(3/2) : (3 : (3/4)) = 3/8$.

Una operación algebraica se denomina *asociativa*, si para cualesquiera tres elementos a, b, c del conjunto inicial se cumple la igualdad $a \circ (b \circ c) = (a \circ b) \circ c$.

La asociatividad de una operación permite hablar sobre el resultado, unívocamente definido, de aplicar la operación algebraica a los tres elementos a, b, c, con la particularidad de que por resul-

lado se obtiene cualquiera de las expresiones iguales $a * (b * c)$ y $(a * b) * c$; además la asociatividad nos permite escribir $a * b * c$ sin paréntesis.

En el caso de una operación asociativa se puede hablar también sobre la univocidad de la expresión $a_1 * a_2 * \ldots * a_n$ que contiene cualquier número finito de elementos $a_1, a_2, \ldots, a_n$. Por significado de $a_1 * a_2 * \ldots * a_n$ entenderemos lo siguiente. En esta expresión distribuyamos los paréntesis de un modo arbitrario, con tal de que se pueda definirla mediante la aplicación sucesiva de una operación algebraica a los pares de elementos. Por ejemplo, para cinco elementos a_1, a_2, a_3, a_4, a_5 los paréntesis podrían ser distribuidos así: $a_1 * ((a_2 * a_3) * (a_4 * a_5))$ o bien: $((a_1 * a_2) * a_3) * (a_4 * a_5)$ o mediante muchos otros métodos.

Demostremos que para una operación asociativa el resultado de cálculo no depende de la distribución de los paréntesis. En efecto, para $n = 3$ esta afirmación se deduce de la definición de operación asociativa. Por ello supongamos $n > 3$ y consideremos que para todos los números inferiores a n nuestra afirmación ya está demostrada.

Sean dados los elementos $a_1, a_2, \ldots, a_n$ y supongamos que los paréntesis están distribuidos de una manera tal que se indica el orden en que debe realizarse la operación. Observemos que el último paso siempre será la realización de la operación con dos elementos, $a_1 * a_2 * \ldots * a_k$ y $a_{k+1} * a_{k+2} * \ldots * a_n$, para cierto k que satisface la condición $1 \leqslant k \leqslant n-1$. Puesto que ambas expresiones contienen menos que n elementos, según la hipótesis ellas se definen unívocamente y nos resta demostrar que para cualesquiera enteros y positivos k, l, $l \geqslant 1$ se verifica la igualdad

$$(a_1 * a_2 * \ldots * a_k) * (a_{k+1} * a_{k+2} * \ldots * a_n) = (a_1 * a_2 * \ldots * a_{k+l}) * (a_{k+l+1} * a_{k+l+2} * \ldots * a_n).$$

Al designar

$$a_1 * a_2 * \ldots * a_k = b,$$

$$a_{k+1} * a_{k+2} * \ldots * a_{k+l} = c,$$

$$a_{k+l+1} * a_{k+l+2} * \ldots * a_n = d,$$

obtenemos, en virtud de la asociatividad de la operación, que

$$b * (c * d) = (b * c) * d,$$

y nuestra afirmación queda demostrada.

Si la operación no es solamente asociativa, sino también conmutativa, entonces la expresión $a_1 * a_2 * \ldots * a_n$ tampoco depende del orden en que se disponen los elementos. La demostración de esta afirmación queda a cargo del lector en calidad de ejercicio.

No conviene pensar que la conmutatividad y la asociatividad de una operación están ligadas entre sí de tal o cual manera. Se pueden construir las operaciones con las más diversas combinaciones de dichas propiedades. Ya hemos visto en los ejemplos de multiplicación y división de los números que una operación puede ser conmutativa y asociativa o bien no conmutativa y no asociativa. Examinemos dos ejemplos más. Supongamos que el conjunto consta de tres elementos a, b, c. Prefijemos las operaciones algebraicas por medio de las tablas:

$\circ$	a	b	c
a	a	c	b
b	c	b	a
c	b	a	c

$*$	a	b	c
a	a	a	a
b	b	b	b
c	c	c	c

(2.1)

y convengamos en que el primero siempre se elige un elemento según la columna, el segundo, un elemento según la fila, mientras que el resultado de la operación se tomará en el lugar donde se intersecan la fila y la columna correspondientes. En el primer caso la operación es, evidentemente, conmutativa, pero no asociativa, puesto que, por ejemplo,

$$(a \circ b) \circ c = c \circ c = c,$$
$$a \circ (b \circ c) = a \circ a = a.$$

En el segundo caso la operación no es conmutativa, sino asociativa, de lo que podemos convencernos con facilidad por una comprobación inmediata.

Ejercicios.

1. ¿Será algebraica la operación de cálculo de $\operatorname{tg} x$ en el conjunto de todos los números x reales?

2. Consideremos un conjunto de números reales x que satisfacen la desigualdad $|x| \leqslant 1$. ¿Serán algebraicas en este conjunto las operaciones de multiplicación, adición, división y sustracción de los números?

3. ¿Será conmutativa y asociativa la operación algebraica $x \circ y = x^2 + y$ en el conjunto de todos los números reales x, y?

4. Supongamos que un conjunto consta de un solo elemento. ¿Cómo se puede definir en este conjunto una operación algebraica?

5. Constrúyanse los ejemplos de operaciones algebraicas en un conjunto cuyos elementos son también conjuntos. ¿Serán estas operaciones conmutativas y asociativas?

§ 3. Operación inversa

Supongamos que en el conjunto A está dada cierta operación algebraica. Como ya sabemos, ésta pone en correspondencia a cualesquiera dos elementos a, b de A un elemento tercero $c = a \circ b$. Consideremos la totalidad C de aquellos

elementos de A que pueden ser representados como resultado de haberse realizado la operación algebraica dada. Por lo visto, cualquiera que sea la operación algebraica, todos los elementos de C son, a la vez, elementos de A. Sin embargo, no es del todo obligatorio que todos los elementos de A pertenezcan a C.

Efectivamente, fijemos en el conjunto A un cierto elemento f y pongámoslo en correspondencia a todo par de elementos a, b de A. Es evidente que la correspondencia construida de tal manera es una operación algebraica y, además, conmutativa y asociativa. El conjunto C sólo contendrá un único elemento f, independientemente de la cantidad de los elementos contenidos en el conjunto A.

Por medio de una operación algebraica se determinan, a veces, precisamente, todos los elementos de A integran C. Sea esta operación tal que C coincide con A, es decir, ambos conjuntos contienen los mismos elementos. En este caso todo elemento de A puede ser representado como resultado de realización de la operación algebraica dada sobre algunos dos elementos del mismo conjunto A. Por supuesto, la representación semejante puede ser no la única. No obstante, llegamos a la conclusión de que a todo elemento de A se le puede poner en correspondencia los pares determinados de elementos de A.

Así pues, la operación algebraica inicial engendra en el conjunto A otra operación. Esta última puede no ser unívoca, puesto que a un solo elemento se le puede poner en correspondencia más que un par de elementos. Pero, incluso siendo unívoca, la operación no será algebraica, dado que está definida no para cualquier *par* de elementos, sino sólo para *un único* elemento, aunque éste último puede ser arbitrario. Con relación a la operación algebraica dada sería natural atribuirle a esta nueva operación el nombre de operación "inversa". Sin embargo, en realidad por operación inversa entenderemos una noción algo diferente, más próxima al concepto de operación algebraica.

Cabe observar que la investigación de la operación "inversa" es equivalente a la investigación de aquellos elementos u, v que satisfacen la igualdad

$$u * v = b \tag{3.1}$$

para diferentes elementos b. La investigación de la ecuación dada respecto de dos elementos u, v se reduce con facilidad a la investigación de dos ecuaciones respecto de un solo elemento. Con este fin basta fijar uno de ellos y determinar el otro, sirviéndose de la ecuación (3.1). Así pues, la investigación de la operación "inversa" es equivalente matemáticamente a la resolución de las ecuaciones

$$a * x = b, \qquad y * a = b \tag{3.2}$$

respecto de los elementos x, y de A para diferentes elementos fijados a, b de A.

Supongamos que las ecuaciones (3.2) tienen soluciones y, además, únicas, para cualesquiera a, b. Entonces, a todo par ordenado de elementos a, b de A podemos ponerle en correspondencia los elementos x, y de A unívocamente definidos, es decir, podemos introducir *dos* operaciones algebraicas. Estas operaciones se llaman operaciones inversas *derecha* e *izquierda*, respectivamente, con relación a la operación principal. Si dichas operaciones existen, diremos que la operación principal tiene operación *inversa*. Observemos que el ejemplo considerado arriba muestra que una operación algebraica, siendo incluso conmutativa y asociativa, puede no tener operaciones inversas, ni la derecha ni la izquierda.

La existencia de la operación inversa significa en realidad que existen, en general, dos operaciones algebraicas diferentes: operaciones inversas derecha e izquierda. Por esta razón hemos de hablar sobre diferentes elementos x, y. En cambio, si una operación algebraica es conmutativa y existe para ella la operación inversa, entonces, evidentemente, $x = y$, y la operación inversa derecha coincide con la izquierda.

He aquí algunos de los ejemplos. Supongamos que el conjunto A representa en sí todo el eje real y la operación *algebraica* consiste en la multiplicación ordinaria de los números. Esta operación en el conjunto *dado* no tiene operación inversa, puesto que, por ejemplo, para $a = 0$, $b = 1$, las igualdades (3.2) no pueden tener lugar, cualesquiera que sean los números x e y. Si, en cambio, examinamos una operación de multiplicación prescrita solamente en el conjunto de números positivos, esta operación, *ahora*, ya tendrá la inversa.

En efecto, para cualesquiera números positivos a y b existen los números positivos x, y —los únicos— que satisfacen las igualdades (3.2). La operación inversa en el caso dado no es otra cosa que división de los números. El hecho de que en realidad $x = y$, no tiene en las condiciones dadas ningún interés para nosotros.

La operación de adición de los números no tiene inversa, si está definida en el conjunto de números *positivos*, puesto que, por ejemplo, las igualdades (3.2) no pueden verificarse, cualesquiera que sean los números positivos x, y, siempre que $a = 2$, $b = 1$. Si, en cambio, la operación de adición de los números está dada en *todo* el eje real, entonces la operación inversa existe y es nada más que la sustracción de números.

El ejemplo con las operaciones de multiplicación y adición de los números muestra que las operaciones directa e inversa pueden poseer unas *propiedades* más diversas. No es del todo obligatorio que de la asociatividad y la conmutatividad de una operación algebraica provenga la asociatividad y la conmutatividad de la operación inversa, incluso si ésta existe. Más aún, como ya se ha señalado más arriba, una operación algebraica conmutativa y asociativa simplemente puede no tener operaciones inversas, ni la derecha ni la izquierda.

Estos ejemplos sencillos indican una circunstancia de importancia más. Examinemos otra vez la operación de multiplicación en un conjunto de números positivos. Para tal operación las operaciones inversas derecha e izquierda coinciden y representan en sí división de los números. Puede parecer a primera vista que ahora para la operación de división de los números la operación inversa consistirá en multiplicación de los números. No obstante, esto no es del todo así.

Efectivamente, escribamos las ecuaciones correspondientes (3.2)

$$a : x = b, \qquad y : a = b.$$

Es evidente que

$$x = a : b, \qquad y = a \cdot b.$$

Por consiguiente, la operación inversa derecha para dividir los números es de nuevo la división de los números, mientras que la operación inversa izquierda es la multiplicación de los números. De este modo, una operación, inversa a la inversa, no coincide necesariamente con la operación algebraica inicial.

Ejercicios.

1. ¿Existirán las operaciones inversas derechas e izquierdas para las operaciones algebraicas definidas por las tablas (2.1)?
2. ¿Qué representan en sí las operaciones inversas derecha e izquierda para la operación algebraica $x \circ y = x^y$, definida en un conjunto de los números positivos x, y?
3. Demuéstrese que si las operaciones inversas derecha e izquierda coinciden, entonces la operación algebraica inicial es conmutativa.
4. Demuéstrese que si una operación algebraica tiene la inversa, entonces las operaciones inversas derecha e izquierda también tienen sus inversas. ¿Qué representan en sí estas operaciones?
5. Constrúyase un ejemplo de la operación algebraica para la cual las cuatro operaciones, inversas a las inversas, son todas coincidentes con la operación inicial.

§ 4. Relación de equivalencia

Observemos que en el examen realizado de las propiedades de una operación algebraica suponíamos implícitamente la posibilidad de que cualesquiera dos elementos del conjunto pueden ser comprobados: si son coincidentes entre sí o no lo son. Más aún, los elementos coincidentes fueron tratados con toda la facilidad sin establecer diferencia entre ellos en ninguno de los casos. Nunca suponíamos que los elementos coincidentes representan, de hecho, un solo elemento y no son los objetos diferentes. En realidad, sólo hemos aprovechado el hecho de que cierto grupo de elementos, que se llamaban iguales, en ciertas condiciones se pone de manifiesto de una manera idéntica.

Con tal situación nos encontramos con frecuencia. Al investigar las propiedades generales de los triángulos semejantes, no hacemos distinciones, en realidad, entre ningunos de los triángulos que tienen ángulos iguales. Desde el punto de vista de las propiedades que se conservan en una transformación de semejanza, los triángulos mencionados no se distinguen y podrían denominarse "iguales". Al investigar los criterios de igualdad de los triángulos, no establecemos diferencia entre aquellos que están dispuestos por diferentes lugares del plano, pero pueden ser hechos coincidentes durante su desplazamiento.

En las más diversas situaciones nos encontraremos con la necesidad de partir tal o cual conjunto en los grupos de elementos unidos entre sí según cierto criterio. Si en estas circunstancias ninguno de los elementos pertenece a los dos grupos diferentes, diremos que se trata de la partición de un conjunto en *grupos disjuntos* o en *clases*.

Los criterios según los cuales los elementos de un conjunto se parten en clases, aunque pueden ser muy distintos, sin embargo, no son del todo arbitrarios. Supongamos, por ejemplo, que deseáramos partir en clases todos los números reales incorporando los números a y b a una misma clase cuando, y sólo cuando, $b > a$. Entonces, ninguno de los números podrá caer a una clase consigo mismo, puesto que a no es mayor que el propio a. Por consiguiente, no puede existir ninguna partición en clases según el criterio dado.

Sea dado un cierto criterio. Convengamos en considerar que respecto a todo par de elementos a, b del conjunto A podemos decir que o bien el elemento a está ligado con el elemento b mediante el criterio o bien no está ligado con él. Si el elemento a está ligado con b, escribiremos $a \sim b$ y diremos que *a es equivalente a b*.

El análisis de los ejemplos más sencillos ya nos dicta aquellas condiciones que deben ser satisfechas por el criterio para que se pueda realizar, sobre la base de éste, una partición del conjunto A en clases. A saber:

1. La reflexividad: $a \sim a$ para todo $a \in A$.
2. La simetría: si $a \sim b$, entonces $b \sim a$.
3. La transitividad: si $a \sim b$ y $b \sim c$, entonces $a \sim c$.

Un criterio que satisface estas condiciones, se denomina *relación de equivalencia*.

Demostraremos que cualquier relación de equivalencia permite partir un conjunto en clases. Efectivamente, sea K_a un grupo de elementos de A equivalentes al elemento fijado a. Debido a la propiedad de reflexividad, $a \in K_a$. Probemos que dos grupos K_a y K_b o bien coinciden o bien no tienen elementos comunes.

Supongamos que cierto elemento c pertenece a K_a y a K_b, es decir, $c \sim a$ y $c \sim b$. En virtud de la propiedad de simetría, $a \sim c$, y, debido a la propiedad de transitividad, $a \sim b$ y, por supuesto, $b \sim a$. Ahora, si $x \in K_a$, entonces $x \sim a$ y, por lo tanto, $x \sim b$,

es decir, $x \in K_b$. Análogamente, si $x \in K_b$ se desprende que $x \in K_a$. De este modo, dos grupos que tienen al menos un solo elemento común, son totalmente coincidentes y hemos obtenido realmente la partición del conjunto A en clases.

Desde el punto de vista del criterio en consideración, cualesquiera dos elementos pueden ser o bien equivalentes o bien no equivalentes. Nada cambiará, si *denominamos* iguales los elementos equivalentes (con relación al criterio dado) y desiguales los no equivalentes (con relación al mismo criterio).

Podría parecer que en este caso despreciamos el sentido de la palabra «iguales», es que ahora unos elementos, iguales según un criterio, puedan resultar *ser* desiguales para otro criterio. No obstante, en esto no hay nada que no sea natural. En todo problema concreto los elementos se distinguen o no se distinguen solamente respecto de aquellas propiedades que son de interés para nosotros precisamente en el problema dado. *Entre tanto*, en los problemas diferentes podemos revelar interés hacia diferentes propiedades de los mismos elementos.

En lo sucesivo consideraremos que siempre, cuando sea necesario, para los elementos de un conjunto debe ser definido un criterio de igualdad que permita decir que el elemento a *es* igual al elemento b o no es igual a éste. Si el elemento a es igual al elemento b, escribiremos $a = b$, y $a \neq b$, en el caso contrario. Supondremos también que el criterio de igualdad es la relación de equivalencia. En este caso las condiciones de reflexividad, simetría y transitividad pueden considerarse como reflejo de las propiedades más generales de una relación ordinaria de igualdad de los números.

La introducción de la relación de igualdad permite partir todo el conjunto en las clases de elementos los cuales, en virtud de una u otra razón, hemos decidido considerar iguales. Quiere decir, la diferencia entre los elementos que integran una misma clase no nos toca de ninguna manera. Por consiguiente, en todas las situaciones que han de ser consideradas en lo sucesivo, los elementos llamados iguales deben ponerse de manifiesto de un modo igual.

Al introducir la relación de igualdad de una manera *axiomática*, es decir, sin referencias a la naturaleza concreta de los elementos, convengamos en considerar que el uso del signo de igualdad sólo significa que los elementos dispuestos por ambos lados de este signo simplemente coinciden: esto es un mismo elemento. Si el signo de igualdad se emplea de la manera indicada, las propiedades de reflexividad, simetría y transitividad no requieren un acuerdo especial. Al partir el conjunto en las clases de elementos iguales, toda clase se compondrá de un único elemento.

En el caso en que la relación de igualdad se introduce basándose sobre *la naturaleza concreta de los elementos*, puede suceder que algunas o incluso todas las clases de elementos iguales se compondrán

do más de un elemento. Esto requiere que al introducir tales u otras operaciones con los elementos, se deben imponer ciertas exigencias adicionales sobre las propias operaciones.

En efecto, según se ha acordado, los elementos iguales deben ponerse de manifiesto de una manera igual. Por esta razón, toda operación que se introduce, siendo aplicada a los elementos iguales, ha de conducir a los resultados iguales. En realidad nunca nos detendremos en la comprobación de que esta exigencia se cumpla, prestando al propio lector la posibilidad de convencerse de la validez de la propiedad citada para todas las operaciones introducidas.

Ejercicios.

1. ¿Pueden partirse todos los países del globo terrestre en clases, incorporando a una misma clase dos países cuando, y sólo cuando, ellos tienen la frontera común? Si no es posible, ¿por qué?

2. Examinemos un conjunto de ciudades que tienen vías de comunicación por automóvil. Llamaremos ligadas dos ciudades A y B, si de A se puede llegar a B por carretera. ¿Se pueden partir las ciudades, según dicho criterio, en clases? Si es posible, ¿qué representan en sí las clases?

3. Llamaremos dos números complejos a y b iguales en módulo, si $|a| = |b|$. ¿Será este criterio la relación de equivalencia? ¿Qué representa en sí la partición en clases?

4. Examinemos las operaciones algebraicas de adición y multiplicación de los números complejos. ¿Cómo operan en las clases de elementos iguales en módulo?

5. Constrúyanse los ejemplos de operaciones algebraicas en el conjunto definido en el ejercicio 2. ¿Cómo estas operaciones actúan en las clases?

§ 5. Segmentos dirigidos

Los ejemplos examinados arriba pueden crear la impresión de que todos los razonamientos acerca de las operaciones sobre elementos de los conjuntos sólo se relacionan con las operaciones sobre diversos conjuntos de números. Sin embargo, esto no es así. En adelante daremos a conocer varios ejemplos de conjuntos no numéricos con operaciones, pero por ahora consideraremos solamente un ejemplo el cual siempre será el objeto de nuestras referencias a lo largo de todo el curso.

Los conceptos más fundamentales de la física son tales nociones como fuerza, desplazamiento, velocidad, aceleración. Todas estas nociones se caracterizan no sólo por un número, que determina su valor, sino también por cierta dirección. Construyamos ahora un análogo geométrico de las nociones semejantes.

Sean dados en el espacio dos puntos distintos A y B. En una línea recta que pasa por los puntos estos últimos determinan de modo natural cierto segmento. Convengamos en considerar que los puntos siempre se dan en un orden determinado, por ejemplo, primero se fija A y después, B. Ahora podremos determinar la dirección

en el segmento construido, a saber, la dirección del primer punto A al segundo punto B.

Un segmento, considerado junto con la dirección prefijada en él, se denomina *segmento dirigido* con el origen A y el extremo B. El segmento dirigido se llamará en otras palabras *vector* y el punto A, *punto de aplicación* del vector. Un vector con el punto de aplicación A lo denominaremos *fijado* en el punto A.

Para los segmentos dirigidos o vectores emplearemos dos formas de su designación. Si hemos de recalcar que se trata de un segmento dirigido cuyo origen se ubica en el punto A y el extremo, en el punto B, escribiremos el símbolo $\overrightarrow{AB}$. Si no nos interesa cuáles precisamente puntos de un segmento dirigido son de frontera, entonces en este caso emplearemos designaciones más sencillas, por ejemplo, las letras latinas minúsculas. En los dibujos los segmentos dirigidos se designarán por medio de las flechas, con la particularidad de que la punta de la flecha siempre se dispondrá en el punto final del segmento.

Fig. 5.1

En todo segmento dirigido es esencial cuál de los puntos frontera constituye su origen y cuál, su extremo. Por esta razón, los segmentos dirigidos $\overrightarrow{AB}$ y $\overrightarrow{BA}$ se consideran diferentes.

Así pues, podemos construir diversos conjuntos cuyos elementos serán segmentos dirigidos. Antes de introducir las operaciones sobre elementos, llegaremos a un acuerdo de qué segmentos dirigidos se considerarán iguales.

Examinemos primero un *traslado paralelo* del segmento dirigido $\overrightarrow{AB}$ en el punto C. Supongamos que el punto C no se dispone en una recta que pasa por A y B (fig. 5.1). Tracemos una recta que pasa por los puntos A y C, después, una recta que pasa por el punto C y es paralela a la recta AB y, por fin, una recta que pasa por el punto B y es paralela a la recta AC. El punto de intersección de las dos últimas rectas se designará mediante D. El segmento dirigido $\overrightarrow{CD}$ se considerará obtenido como resultado de trasladar paralelamente el segmento $\overrightarrow{AB}$ en el punto C. Si, en cambio, el punto C se ubica en la recta que pasa por A y B, entonces el segmento $\overrightarrow{CD}$ se obtiene por el desplazamiento del segmento $\overrightarrow{AB}$ a lo largo de la recta que lo contiene hasta que el punto A coincida con el punto C.

Ahora se puede enunciar la definición de igualdad de los vectores. Dos vectores se denominan *iguales*, si se pueden hacer coincidir en

el transcurso del traslado paralelo. No es difícil ver que esta definición de igualdad es la relación de equivalencia, es decir, posee las propiedades de reflexividad, simetría y transitividad.

De este modo, la totalidad de todos los vectores se divide de modo natural en unas clases de vectores iguales. Cada una de estas clases se describe sin dificultades. Se obtiene por medio del traslado paralelo de cualquiera de los vectores de la clase en todos los puntos del espacio.

Observamos que en todo punto del espacio se halla fijado un vector y sólo uno, de cada clase de los vectores iguales. Por esto, al comparar los vectores a, b se puede usar el siguiente procedimiento. Habiendo fijado un cierto punto, trasladamos paralelamente en este punto los vectores a, b. Si, en este caso, ellos coinciden por completo, entonces $a = b$, y $a \neq b$, en el caso contrario.

Además de un conjunto compuesto por todos los vectores del espacio, trataremos con frecuencia unos conjuntos de otro género. En lo principal, serán los conjuntos de vectores o bien paralelos a cierta línea recta o dispuestos en ésta, o bien paralelos a cierto plano o dispuestos en éste. Los vectores de tal índole los llamaremos *colineales* o *coplanares*, respectivamente. Naturalmente en los conjuntos de vectores colineales y coplanares sigue siendo válida la definición de igualdad de los vectores enunciada más arriba.

Consideraremos también los así llamados segmentos dirigidos *nulos* cuyos orígenes coinciden con los extremos. La orientación de los vectores nulos no está definida y todos ellos se consideran, por definición, iguales. Si la indicación de los puntos frontera de un vector nulo no es obligatoria, dicho vector se designará por el símbolo $\mathbf{0}$.

También por definición, consideraremos que cualquier vector nulo es paralelo a toda recta y a todo plano. Por esta razón, siempre supondremos en lo sucesivo, si no se hace reserva especial, que el conjunto de vectores de un espacio, como también todo conjunto de los vectores colineales y coplanares comprende en sí el conjunto de todos los vectores nulos. Esto no se debe olvidar.

Ejercicios.

1. Demuéstrese que los vectores no nulos de un espacio pueden ser divididos en clases de vectores colineales no nulos.
2. Demuéstrese que cualquier clase de vectores colineales no nulos se la puede dividir en clases de vectores iguales no nulos.
3. Demuéstrese que toda clase de vectores iguales no nulos se incorpora íntegramente a una y sólo una clase de vectores colineales no nulos.
4. ¿Pueden dividirse los vectores no nulos de un espacio en las clases de vectores coplanares? ¿Si no se dividen, por qué?
5. Demuéstrese que todo conjunto de vectores coplanares no nulos se lo puede dividir en las clases de vectores colineales no nulos.

6. Demuéstrese que todo par de diferentes clases de los vectores colineales no nulos se incorpora íntegramente a uno y sólo un conjunto de vectores coplanares no nulos.

§ 6. Adición de los segmentos dirigidos

Como ya se ha indicado, la fuerza, el desplazamiento, la velocidad y la aceleración son imágenes de los segmentos dirigidos construidos por nosotros. Para que estos segmentos resulten útiles en la resolución de diferentes problemas físicos, se debe, al introducir las operaciones con ellos, tomar en consideración también las analogías físicas correspondientes.

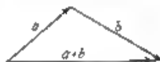


Fig. 6.1.

Está bien conocida la operación de adición de fuerzas que se realiza según la así llamada regla del paralelogramo. Conforme a esta misma regla se suman también los desplazamientos, las velocidades y las aceleraciones. De acuerdo con la terminología introducida, esta operación es algebraica, conmutativa y asociativa. Nuestra tarea inmediata consiste en construcción de una operación semejante sobre los segmentos dirigidos.

Definamos la operación de *adición de los vectores* del modo siguiente. Supongamos que es preciso sumar los vectores a y b. Traslademos paralelamente el vector b en el extremo del vector a (fig. 6.1). Entonces, se llamará *suma* $a + b$ un vector cuyo origen coincide con el origen de a y el extremo, con el extremo de b. Esta regla para realizar la operación se denomina, comúnmente, «*regla del triángulo*».

Es evidente que la operación de adición de los vectores es algebraica. Demostremos que es conmutativa y asociativa.

Para demostrar la conmutatividad de la operación supongamos al principio que los vectores a y b no son colineales. Dispongámoslos en el origen general O (fig. 6.2). Designaremos con las letras A y B los extremos de los vectores a y b, respectivamente y consideremos el paralelogramo $OBCA$. De la definición de igualdad de los vectores se deduce que

$$\overrightarrow{BC} = \overrightarrow{OA} = a, \qquad \overrightarrow{AC} = \overrightarrow{OB} = b.$$

Pero, en este caso, es la misma diagonal $\overrightarrow{OC}$ del paralelogramo $OBCA$ es, a la vez, $a + b$ y $b + a$. El caso colineal de los vectores a y b es obvio.

Observemos que al mismo tiempo hemos obtenido otro método para construir la suma de los vectores. A saber, si en los vectores a, b fijados en un punto, está construido un paralelogramo, su diagonal fijada en el mismo punto será la suma $a + b$.

Con ol objeto do demostrar la asociatividad de la operación de adición, aplicamos el vector a al punto arbitrario O, el vector b, al extremo del vector a y el vector c, al extremo dol vector b (fig. 6.3). Designemos con A, B, C los extremos de loo vectores a, b, c. Entonces

$$(a+b)+c=(\overrightarrow{OA}+\overrightarrow{AB})+\overrightarrow{BC}=\overrightarrow{OB}+\overrightarrow{BC}=\overrightarrow{OC},$$

$$a+(b+c)=\overrightarrow{OA}+(\overrightarrow{AB}+\overrightarrow{BC})=\overrightarrow{OA}+\overrightarrow{AC}=\overrightarrow{OC}.$$

Por ser transitiva la relación de igualdad do los vectores concluimos que la asociatividad de la operación también tiene lugar.



Fig. 6.2.

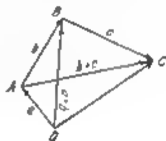


Fig. 6.3.

Las propiedades demostradas de la operación de adición de vectores permiten calcular la suma de cualquier número de vectores. Si el vector a_2 se aplica al extremo del vector a_1, el vector a_3, al extremo del vector a_2, etc. y, por último, el vector a_n, al extremo del vector a_{n-1}, entonces la suma $a_1+a_2+\ldots+a_n$ representará un vector cuyo origen coincide con el origen do a_1 y el extremo, con el de a_n. Esta regla para construir las sumas de vectores se denomina *«regla de cierre de una quebrada hasta completar un polígono»*.

Planteemos ahora una pregunta sobre la existencia de la operación inversa para sumar los vectores. Como es sabido, para contestarla es preciso investigar la existencia y unicidad de la solución de la ecuación

$$a+x=b, \qquad y+a=b$$

para los vectores a, b arbitrarios. En virtud de que la operación principal es conmutativa, será suficiente, evidentemente, investigar sólo una de las ecuaciones.

Tomemos un segmento dirigido arbitrario $\overrightarrow{AB}$. Mediante una construcción geométrica elemental establecemos que siempre se verifican las correlaciones

$$\overrightarrow{AB}+\overrightarrow{BA}=0, \qquad \overrightarrow{AB}+0=\overrightarrow{AB}. \qquad (6.1)$$

Por ello, la ecuación

$$\overrightarrow{AB} + x = \overrightarrow{CD}, \qquad (6.2)$$

para cualesquiera vectores $\overrightarrow{AB}$ y $\overrightarrow{CD}$, tendrá, a ciencia cierta, por lo menos una solución, por ejemplo,

$$x = \overrightarrow{BA} + \overrightarrow{CD}. \qquad (6.3)$$

Supongamos que la ecuación (6.2) queda satisfecha, además, por cierto otro vector z, es decir,

$$\overrightarrow{AB} + x = \overrightarrow{CD}, \qquad \overrightarrow{AB} + z = \overrightarrow{CD}.$$

Entonces, sumando $\overrightarrow{BA}$ a ambos miembros de estas igualdades, obtendremos, al tomar en consideración (6.1), que $x = \overrightarrow{BA} + \overrightarrow{CD}$, $z = \overrightarrow{BA} + \overrightarrow{CD}$ y, por consiguiente, $x = z$.

De este modo, para la operación de adición introducida por nosotros existe una operación inversa. Esta se denomina operación de *sustracción de vectores*. Si para los vectores a, b, c tiene lugar la igualdad $a + c = b$, *escribiremos simbólicamente* que $c = b - a$. El vector $b - a$, definido unívocamente por los vectores b y a, se llama *diferencia* de vectores. Un poco más tarde daremos la argumentación para tal designación y tal nombre.



Fig. 6.4

Es fácil indicar la regla para construir la diferencia entre dos vectores dados a y b. Apliquemos estos vectores a un punto común y construyamos en ellos un paralelogramo (fig. 6.4). Ya hemos señalado anteriormente que una diagonal del paralelogramo es la suma de los vectores dados. La otra diagonal, como es fácil de ver, es la diferencia de los mismos vectores. La regla descrita para construir la suma y la diferencia de los vectores se llama, comúnmente, "*regla del paralelogramo*".

Observemos que podríamos definir la operación de adición no para el conjunto de todos los vectores de un espacio, sino sólo para uno de los conjuntos de vectores colineales o coplanares. La suma de dos vectores de cualquier conjunto de tal tipo pertenecerá de nuevo al *mismo* conjunto. Es por esto que la operación de adición de vectores queda algebraica en este caso también. Más aún, dicha operación sigue conservando todas sus propiedades y, lo que es de mayor importancia, posee, como antes, la operación inversa. La validez de esta última afirmación se desprende de la fórmula (6.3). Si los vectores

$\overrightarrow{AB}$ y $\overrightarrow{CD}$ son paralelos a cierta recta o a cierto plano, entonces es evidente que será paralelo también el vector $\overrightarrow{BA} + \overrightarrow{CD}$ o bien, que es lo mismo, el vector de la diferencia $\overrightarrow{CD} - \overrightarrow{AB}$.

Así pues, la operación de adición de los vectores es algebraica, conmutativa, asociativa y tiene operación inversa en los conjuntos de tres tipos. Estos son: el conjunto de vectores de un espacio, el conjunto de vectores colineales y el conjunto de vectores coplanares.

Ejercicios.

1. A uno de los vértices de un cubo están aplicadas tres fuerzas iguales en valor y dirigidas a lo largo de las aristas. ¿Cuál es la dirección de la suma de estas fuerzas?
2. Sean dadas tres clases diferentes de los vectores colineales. ¿En qué caso cualquier vector del espacio puede ser representado como suma de tres vectores de estas clases?
3. A los vértices de un polígono regular están aplicadas las fuerzas iguales por su valor y dirigidas hacia el centro. ¿Cuál será la suma de estas fuerzas?
4. ¿Qué representa en sí el conjunto de sumas de los vectores tomados de dos clases diferentes de los vectores colineales?

§ 7. Grupos

Los conjuntos con una operación algebraica son en cierto sentido los más sencillos, razón por la cual parece natural empezar nuestras investigaciones precisamente por tales conjuntos. Tomaremos por axiomas las propiedades de la operación y después deduciremos de éstos los corolarios. Esto nos permite *en lo sucesivo* aplicar inmediatamente los resultados de nuestras investigaciones a todos los conjuntos, en los cuales las operaciones tienen propiedades análogas, independientemente de las peculiaridades concretas.

Se denomina *grupo* un conjunto G con una operación algebraica que es asociativa (aunque no sea necesariamente conmutativa), con la particularidad de que para dicha operación debe existir operación inversa.

Ha de notarse que la operación inversa no puede considerarse como segunda operación independiente en el grupo, puesto que ella se determina en términos de la operación principal. Según lo aceptado en la teoría de los grupos, llamemos *multiplicación* la operación prefijada en G y convengamos en emplear las notaciones correspondientes. Antes de empezar a considerar los ejemplos diferentes de grupos, deduzcamos unos corolarios más sencillos de la propia definición.

Examinemos un elemento arbitrario a del grupo G. Del hecho de que en el grupo existe la operación inversa se desprende la existen-

cia de un único elemento e_a de tal índole que $ae_a = a$. Por consiguiente, cuando el elemento a se multiplica a la derecha por el e_a, este último desempeña el mismo papel que en la multiplicación de los números pertenece a la unidad. Ahora, sea b cualquier otro elemento del grupo. Es evidente que existe un elemento y que satisface la igualdad $ya = b$. Obtenemos pues

$$b = ya = y(ae_a) = (ya)e_a = be_a.$$

Así, e_a desempeña el papel de la unidad derecha respecto a todos los elementos del grupo G, y no sólo respecto de a. El elemento que posee la propiedad semejante debe ser único. Efectivamente, todos los elementos de esta índole satisfacen la ecuación $ax = a$, pero, conforme a la definición de operación inversa, dicha ecuación tiene una única solución. Designaremos el elemento obtenido mediante e'.

Análogamente se puede demostrar la existencia y unicidad en el grupo G del elemento e'' que satisface la igualdad $e''b = b$ para todo elemento b de G. En realidad los elementos e' y e'' coinciden, lo que se deduce de las igualdades $e''e' = e''$ y $e''e' = e'$.

De este modo hemos obtenido el primer corolario importante: en cualquier grupo G existe un elemento e y, además, *único* que satisface las igualdades

$$ae = ea = a$$

para todo a de G. Este elemento recibe el nombre de elemento *unidad* del grupo G.

De la definición de operación inversa se desprende también la existencia y unicidad, para todo elemento a, de tales elementos a' y a'' que

$$aa' = e, \qquad a''a = e.$$

Estos se denominan *elementos inversos derecho e izquierdo*, respectivamente. Es fácil mostrar que en el caso dado ellos coinciden. En efecto, examinemos el elemento $a''aa'$ y calculémoslo empleando dos procedimientos diferentes. Se tiene

$$a''aa' = a''(aa') = a''e = a'',$$
$$a''aa' = (a''a)a' = ea' = a'.$$

Por consiguiente, $a'' = a'$. Este elemento se llama *inverso* al elemento a y se designa mediante a^{-1}.

Hemos obtenido, pues, el segundo corolario de importancia: en todo grupo G cualquier elemento a posee un *único* elemento inverso a^{-1}, para el cual

$$aa^{-1} = a^{-1}a = e. \qquad (7.1)$$

Basándose en la asociatividad de la operación de grupo se puede hablar sobre la univocidad del producto de cualquier número finito

de elementos del grupo prefijado (al tener en cuenta la no conmutatividad eventual de la operación de grupo) en el orden determinado. Tomando en consideración las correlaciones (7.1), no es difícil señalar la fórmula general para un elemento inverso al producto. A saber,

$$(a_1a_2\ldots a_n)^{-1}=a_n^{-1}a_{n-1}^{-1}\ldots a_1^{-1}. \qquad (7.2)$$

De la correlación (7.1) se deduce que de elemento inverso respecto de a^{-1} sirve el propio elemento a, y el elemento inverso para la unidad será la propia unidad, es decir

$$(a^{-1})^{-1}=a, \qquad e^{-1}=e. \qquad (7.3)$$

La comprobación de que si es o no un grupo, el conjunto con una operación asociativa, se facilita mucho gracias al hecho de que en la definición de grupo el requisito de que se cumpla la operación inversa puede sustituirse por la suposición de que existen la unidad y los elementos inversos y existen, además, sólo por un lado (a la derecha, por ejemplo) sin que se suponga la unicidad de la unidad y de los elementos citados. Para precisar, resulta válido el siguiente

TEOREMA 7.1. *El conjunto G con una operación asociativa será un grupo, si en G existe al menos un elemento e que posee la propiedad de que $ae=a$ para todo a perteneciente a G, y respecto de éste todo elemento a de G tiene aunque un elemento inverso derecho a^{-1}, es decir, $aa^{-1}=e$.*

DEMOSTRACIÓN. Sea a^{-1} uno de los elementos inversos derechos para a. Se tiene

$$aa^{-1}=e=ee=eaa^{-1}.$$

Multipliquemos ambos miembros de esta igualdad a la derecha por uno de los elementos que son derechos e inversos para a^{-1}. Obtendremos que $ae=eae$, de donde se desprende la igualdad $a=ea$, puesto que e es la unidad derecha para G. De este modo, el elemento e resulta ser también la unidad izquierda para G.

Luego, si e' es una unidad arbitraria derecha y e'', una unidad arbitraria izquierda, entonces de las igualdades $e''e'=e''$ y $e''e'=e'$ proviene que $e'=e''$, es decir, toda unidad derecha equivale a cualquiera izquierda. Con esto quedan demostradas la existencia y unicidad en el conjunto G del elemento unidad al cual se designará de nuevo mediante e.

Ahora, para todo elemento inverso derecho a^{-1} se verifican las siguientes correlaciones:

$$a^{-1}=a^{-1}e=a^{-1}aa^{-1}.$$

Multipliquemos ambos miembros de esta igualdad a la derecha por uno de los elementos, derechos e inversos para a^{-1}. Llegamos a que $e=a^{-1}a$, es decir, el elemento a^{-1} es, a la vez, elemento inver-

so izquierdo para a. Si, ahora, $a^{-1''}$ es un elemento inverso derecho arbitrario para a y $a^{-1'}$, un elemento inverso izquierdo arbitrario, entonces de las igualdades

$$a^{-1''}aa^{-1'} = (a^{-1''}a)\,a^{-1'} = ea^{-1'} = a^{-1'},$$
$$a^{-1''}aa^{-1'} = a^{-1''}\,(aa^{-1'}) = a^{-1''}e = a^{-1''}$$

se deduce que $a^{-1'} = a^{-1''}$. Esto significa la existencia y unicidad, para todo elemento a de G, de un elemento inverso a^{-1}.

Ahora no es difícil de probar que el conjunto G será un grupo. En efecto, las ecuaciones $ax = b$, $ya = b$ serán, a ciencia cierta, satisfechas por los elementos

$$x = a^{-1}b, \qquad y = ba^{-1}.$$

Supongamos que existen también otras soluciones, por ejemplo, el elemento x_0 para la primera ecuación. En este caso de las igualdades $ax = b$ y $ax_0 = b$ se desprende que $ax = ax_0$. Al multiplicar ambos miembros a la izquierda por el elemento a^{-1}, obtenemos $x = x_0$. De este modo, el conjunto G es un grupo.

Un grupo se llama *conmutativo* o *abeliano*, si la operación de grupo es conmutativa. En este caso la operación, como regla, recibe el nombre de *adición* y en lugar del símbolo de multiplicación ab suele escribirse el de la suma $a + b$. La unidad de un grupo abeliano se llama *elemento nulo* y se designa mediante el símbolo 0. La operación inversa se denomina *sustracción* y el elemento inverso, *opuesto*. Se designa el último por el símbolo $-a$. Convengamos en considerar que, por definición, el símbolo de la diferencia $a - b$ significa la suma $a + (-b)$.

Si, no obstante, en virtud de tal o cual razón, la operación en un grupo conmutativo se llamará *multiplicación*, entonces la operación inversa se considerará como *división*. Los productos $a^{-1}b$ y ba^{-1}, que en el caso dado son iguales, se designarán mediante b/a y se llamarán cociente de la división de b por a.

Ejercicios.

Demuéstrese que los conjuntos a seguir son grupos abelianos. La denominación de toda operación refleja siempre no la notación, sino el contenido.

1. El conjunto: números enteros; la operación: adición de los números.
2. El conjunto: números complejos, salvo cero; la operación: multiplicación de los números.
3. El conjunto: números enteros múltiplos del número 3; la operación: sumación de los números.
4. El conjunto: números enteros racionales; la operación: multiplicación de los números.
5. El conjunto: números del tipo $a + b\sqrt{2}$, donde a, b son números racionales positivos; la operación: multiplicación de los números.
6. El conjunto: un elemento a; la operación se denomina adición y se define mediante la igualdad $a + a = a$.

7. El conjunto: números enteros $0, 1, 2, \ldots, n-1$; la operación se llama *adición según módulo n* y consiste en el cálculo del resto no negativo, inferior a n, de la división de una suma de dos números por el número n.

8. El conjunto: números enteros $1, 2, 3, \ldots, n-1$, donde n es un número primo; la operación se llama *multiplicación según el módulo n* y consiste en el cálculo del resto no negativo, inferior a n, de la división de un producto de dos números por el número n.

9. El conjunto: segmentos dirigidos colineales; la operación: sumación de los segmentos dirigidos.

10. El conjunto: segmentos dirigidos coplanares; la operación: sumación de los segmentos dirigidos.

11. El conjunto: segmentos dirigidos de un espacio; la operación: sumación de los segmentos dirigidos.

En lo que se refiere a los tres últimos ejemplos hemos de notar que el elemento nulo del grupo abeliano de segmentos dirigidos lo constituirá el segmento dirigido nulo y el segmento opuesto para $\overrightarrow{AB}$ será el segmento $\overrightarrow{BA}$. De lo demostrado más arriba proviene su unicidad. Los ejemplos de grupos no conmutativos los aduciremos más tarde.

§ 8. Anillos y campos

Examinemos un conjunto K en el que se han introducido dos operaciones. Se llamará adición una de las operaciones y la otra será multiplicación; se empleará la notación correspondiente. Supondremos que ambas operaciones están correlazadas mediante la ley *distributiva*, es decir, para cualesquiera tres elementos a, b, c de K tienen lugar las correlaciones

$$(a+b)\,c = ac + bc, \qquad a\,(b+c) = ab + ac.$$

El conjunto K se llama *anillo*, si en él están definidas dos operaciones: adición y multiplicación, siendo asociativas ambas y ligadas entre sí por la ley distributiva, con la particularidad de que la adición es conmutativa y posee operación inversa. El anillo se denomina *conmutativo*, si la multiplicación es conmutativa, y *no conmutativo*, en el caso contrario.

Observemos que cualquier anillo es un grupo abeliano por adición. Por consiguiente, está provisto de un único elemento nulo 0. Dicho elemento tiene la propiedad de que para todo elemento a del anillo se verifica la igualdad

$$a + 0 = a.$$

La definición de elemento nulo se ha proporcionado únicamente con relación a la operación de adición. No obstante, respecto a la multiplicación este elemento también desempeña un papel singular. A saber, en todo anillo el producto de cualquier elemento por el elemento nulo es elemento nulo. Efectivamente, sea a un elemento cualquiera del anillo K, entonces

$$a\cdot 0 = a\,(0+0) = a\cdot 0 + a\cdot 0.$$

Sumando a ambos miembros de esta igualdad el elemento $-a\cdot 0$, llegamos a que $a\cdot 0 = 0$. Análogamente se demuestra que también $0\cdot a = 0$.

Al hacer uso de esta propiedad del elemento nulo podemos demostrar que en todo anillo para cualesquiera elementos a, b es válida la igualdad

$$(-a)\, b = -(ab).$$

En efecto,

$$ab + (-a)\, b = (a + (-a))\, b = 0\cdot b = 0,$$

es decir, el elemento $(-a)\, b$ es opuesto para el elemento ab. De conformidad con nuestras designaciones podemos escribirlo en la forma $-(ab)$.

Ahora es fácil probar que la ley distributiva es lícita también para la diferencia de elementos. Tenemos

$$(a - b)\, c = (a + (-b))\, c = ac + (-b)\, c = $$

$$= ac + (-(bc)) = ac - bc,$$

$$a\, (b - c) = a\, (b + (-c)) = ab + a\, (-c) =$$

$$= ab + (-(ac)) = ab - ac.$$

La ley distributiva, es decir, la regla común para abrir los paréntesis, constituye una única exigencia en la definición de anillo que liga la adición y la multiplicación. Solamente gracias a esta ley el estudio conjunto de dos operaciones citadas nos suministra más que se podría obtener al estudiarlas por separado.

Acabamos de demostrar que las operaciones algebraicas en un anillo poseen muchas propiedades habituales para las operaciones sobre los números. No conviene pensar, sin embargo, que cualquier propiedad de la adición y multiplicación de los números queda en vigor para todo anillo, aunque sea este último conmutativo. Así, por ejemplo, la multiplicación de los números posee una propiedad que es inversa a la de multiplicación por un elemento nulo. A saber si un producto de dos números es nulo, entonces por lo menos uno de los factores es igual a cero. Esta propiedad ya no es obligatoria para un anillo conmutativo, es decir, un producto de los elementos, distintos de elemento nulo, puede resultar nulo.

Los elementos no nulos cuyo producto es un elemento nulo se denominan *divisores de cero*. La presencia de tales elementos en el anillo hace la investigación mucho más complicada y no presta la oportunidad de establecer analogía profunda entre los números y los elementos del anillo conmutativo. La analogía de este género nos la proporciona el examen de los anillos privados de divisores de cero.

Supongamos que con relación a la operación de multiplicación en un anillo conmutativo hay un elemento unidad e y todo elemento no nulo a tiene elemento inverso a^{-1}. No es difícil demostrar

que tanto elemento unidad como el inverso son únicos y, no obstante, lo más importante consiste en el hecho de que ahora el anillo no tiene divisores de cero. Efectivamente, sea $ab = 0$, pero $a \neq 0$. Al multiplicar ambos miembros de esta igualdad a la izquierda por el elemento a^{-1}, llegamos a que

$$a^{-1}ab = (a^{-1}a)\,b = eb = b$$

y, por supuesto, $a^{-1}0 = 0$. Por consiguiente, $b = 0$.

De la ausencia de los divisores de cero se deduce que toda igualdad se puede reducir por un factor común no nulo. Si es que $ca = cb$ y $c \neq 0$, entonces $c\,(a - b) = 0$, de lo cual deducimos que $a - b = 0$, es decir, $a = b$.

Se denomina *campo**) un anillo conmutativo P que contiene un elemento unidad y todo elemento no nulo tiene elemento inverso.

Aprovechando la inscripción de un cociente a/b en forma de un producto ab^{-1}, podemos probar con facilidad que *en cada campo quedan en vigor todas las reglas ordinarias para operar con fracciones, teniendo en cuenta las operaciones de adición, sustracción, división y multiplicación.* A saber:

$$\frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd}, \qquad \frac{a}{b}\cdot\frac{c}{d} = \frac{ac}{bd}, \qquad \frac{-a}{b} = -\frac{a}{b}.$$

Además, $a/b = c/d$ cuando, y sólo cuando, $ad = bc$, si, naturalmente, $b \neq 0$ y $d \neq 0$. Queda a cargo del lector comprobar la validez de estas afirmaciones en calidad de los ejercicios.

Así, desde el punto de vista de las reglas ordinarias para operar con fracciones, no hay campo que se diferencie de un conjunto de números, razón por la cual los elementos *de todo* campo los llamaremos números, siempre que, por supuesto, tal denominación no nos lleve a una ambigüedad. Como regla, el elemento nulo de cualquier campo se designará por el símbolo 0, y el elemento unidad, por el símbolo 1.

Enunciaremos ahora todos los datos comunes sobre los elementos de cualquier campo conmutativo que nos harán falta en lo sucesivo.

A. A todo par de elementos a, b le corresponde un elemento $a + b$, llamado suma de a y b, con la particularidad de que:

1) la adición es conmutativa, $a + b = b + a$,

2) la adición es asociativa, $a + (b + c) = (a + b) + c$,

3) existe un único elemento nulo 0 tal que $a + 0 = a$ para cualquier elemento a,

4) para todo elemento a existe un único elemento opuesto $-a$ tal que $a + (-a) = 0$.

*) Al igual que en la obra de Kurosch A. G. «Curso de álgebra superior», editorial «MIR», Moscú, consideramos conveniente, para brevedad y comodidad, emplear el término «campo» como un sinónimo de «cuerpo conmutativo». (*N. del Tr.*)

B. A todo par de elementos a, b le corresponde un elemento ab, llamado producto de a y b, con la particularidad de que:

1) la multiplicación es conmutativa, $ab = ba$,

2) la multiplicación es asociativa, $a(bc) = (ab)c$,

3) existe un único elemento unidad 1 tal que $a \cdot 1 = 1 \cdot a = a$ para todo elemento a,

4) para todo elemento no nulo a existe un único elemento inverso a^{-1} tal que $aa^{-1} = a^{-1}a = 1$.

C. Las operaciones de adición y multiplicación están ligadas entre sí mediante la siguiente correlación: la multiplicación distributiva respecto de la adición, $(a + b)c = ac + bc$.

Los datos citados no pretenden constituir una independencia lógica, sino que sólo son una característica cómoda de los elementos. Las propiedades A describen un campo desde el punto de vista de la operación de adición y dicen que respecto a dicha operación el campo constituye un grupo abeliano. Las propiedades B describen un campo desde el punto de vista de la operación de multiplicación e indican que respecto a dicha operación el campo se hace un grupo abeliano, si del campo se excluye el elemento nulo. La propiedad C describe los lazos que unen entre sí dos operaciones.

Ejercicios.

Demuéstrese que los conjuntos 1—7 son los anillos, pero no los campos, mientras que los conjuntos 8—13 son los campos. La denominación de la operación refleja en todos los casos no la notación, sino el contenido.

1. El conjunto: números enteros; las operaciones: adición y multiplicación de los números.

2. El conjunto: números enteros múltiplos de cierto número n; las operaciones: adición y multiplicación de los números.

3. El conjunto: números reales del tipo $a + b\sqrt{2}$, donde a y b son números enteros; las operaciones: adición y multiplicación de los números.

4. El conjunto: polinomios con coeficientes reales de una variable t, incluidas las constantes; las operaciones: adición y multiplicación de los polinomios.

5. El conjunto: un solo elemento a; las operaciones se definen por las igualdades $a + a = a$ y $a \cdot a = a$.

6. El conjunto: números enteros $0, 1, 2, \ldots, n - 1$, donde n es un número compuesto; las operaciones: adición y multiplicación según el módulo n.

7. El conjunto: pares (a, b) de números enteros; las operaciones se definen mediante las fórmulas

$$(a, b) + (c, d) = (a + c, b + d); \quad (a, b) \cdot (c, d) = (ac, bd).$$

8. El conjunto: números racionales; las operaciones: adición y multiplicación de los números.

9. El conjunto: números reales; las operaciones: adición y multiplicación de los números.

10. El conjunto: números complejos; las operaciones: adición y multiplicación de los números.

11. El conjunto: números reales del tipo $a + b\sqrt{2}$, donde a y b son números racionales; las operaciones: adición y multiplicación de los números.

12. El conjunto: dos elementos a y b; las operaciones se definen por las igualdades

$$a + a = b + b = a, \qquad a + b = b + a = b,$$
$$a \cdot a = a \cdot b = b \cdot a = a, \qquad b \cdot b = b.$$

13. El conjunto: números enteros $0, 1, 2, \ldots, n-1$, donde n es un número primo; las operaciones: adición y multiplicación según el módulo n.

Llamamos la atención del lector a que en uno de los ejemplos se ha aducido un anillo con divisores de cero. ¿Cuál es el ejemplo? ¿Cuál es la forma general de los divisores de cero?

§ 9. Multiplicación del segmento dirigido por un número

Recalquemos una vez más que la operación algebraica se ha definido como operación sobre dos elementos pertenecientes a un mismo conjunto. Sin embargo, los ejemplos numerosos de la física muestran que a veces resulta razonable también considerar operaciones sobre los elementos pertenecientes a diversos conjuntos. Una de tales operaciones es dictada por los conceptos de fuerza, desplazamiento, velocidad y aceleración; considerémosla recurriendo de nuevo a los ejemplos de segmentos dirigidos.

En la física ya hace tiempo se aceptó la práctica de utilizar los segmentos. Si se dice que, por ejemplo, una fuerza se ha aumentado cinco veces, el segmento que la representa se hace "extender" cinco veces sin cambiar la dirección general. Si se habla sobre el cambio de la dirección en que actúa la fuerza, entonces en el segmento correspondiente los puntos inicial (origen) y final cambian de lugar. Partiendo de estos razonamientos, introduzcamos una operación que se denomina multiplicación del segmento dirigido por un número real.

Consideraremos al principio algunas cuestiones generales. Sea dada en un plano o en un espacio una línea recta arbitraria. Convengamos en tomar una de las direcciones en ésta por positiva y la contraria, por negativa. Una recta con la dirección definida se llamará *eje*.

Supongamos ahora que se ha dado un eje e indicado, además, el segmento graduado con ayuda del cual podamos medir cualquier otro segmento y determinar, consecuentemente, la longitud del último. A todo segmento dirigido se le atribuirá su característica numérica, la así llamada magnitud del segmento dirigido.

Se denomina *magnitud* $(\overrightarrow{AB})$ del segmento dirigido $\overrightarrow{AB}$ un número, igual a la longitud del segmento $\overrightarrow{AB}$, que se toma con el signo "más", si la dirección de $\overrightarrow{AB}$ coincide con la dirección positiva del eje y con el signo "menos", si la dirección de $\overrightarrow{AB}$ coincide con la

dirección negativa del eje. Las magnitudes de todos los segmentos dirigidos nulos se consideran iguales a cero, es decir,

$$(\overrightarrow{AA}) = 0.$$

Independientemente de cuál dirección en el eje se ha aceptado por positiva, la dirección de $\overrightarrow{AB}$ es opuesta a la de $\overrightarrow{BA}$, mientras que las longitudes de los segmentos dirigidos $\overrightarrow{AB}$ y $\overrightarrow{BA}$ son iguales, por lo cual,

$$(\overrightarrow{AB}) = -(\overrightarrow{BA}). \qquad (9.1)$$

La magnitud de un segmento dirigido, a diferencia de su longitud, puede tener cualquier signo. Puesto que la longitud del segmento



Fig. 9.1.

dirigido $\overrightarrow{AB}$ es el módulo de su magnitud, entonces con el fin de designarla emplearemos el símbolo $|\overrightarrow{AB}|$. Está claro que a diferencia de (9.1)

$$|\overrightarrow{AB}| = |\overrightarrow{BA}|.$$

Supongamos que en un eje están prefijados tres puntos cualesquiera A, B, C que definen tres segmentos dirigidos $\overrightarrow{AB}$, $\overrightarrow{BC}$ y $\overrightarrow{AC}$. Cualquiera que sea la disposición de los puntos, las magnitudes de dichos segmentos dirigidos satisfacen la correlación

$$(\overrightarrow{AB}) + (\overrightarrow{BC}) = (\overrightarrow{AC}). \qquad (9.2)$$

En efecto, supongamos que la dirección del eje y la disposición de los puntos es tal como las muestra la fig. 9.1. En este caso será evidente que

$$|\overrightarrow{CA}| + |\overrightarrow{AB}| = |\overrightarrow{CB}|. \qquad (9.3)$$

De acuerdo con la definición de magnitud del segmento dirigido y con la igualdad (9.1) tenemos

$$|\overrightarrow{CA}| = (\overrightarrow{CA}) = -(\overrightarrow{AC}), \qquad |\overrightarrow{AB}| = (\overrightarrow{AB}), \qquad (9.4)$$

$$|\overrightarrow{CB}| = (\overrightarrow{CB}) = -(\overrightarrow{BC}).$$

Por ello, de (9.3) se deduce

$$-(\overrightarrow{AC})+(\overrightarrow{AB})=-(\overrightarrow{BC}),$$

lo que coincide, en esencia, con (9.2).

En la demostración hemos usado únicamente las correlaciones (9.3) y (9.4) que sólo dependen de la disposición mutua de los puntos A, B, C en el eje, pero no dependen de si dichos puntos coinciden o no uno con el otro. Está claro que con cualquier otra disposición de los puntos la demostración se realiza del modo análogo.

La identidad (9.2) se llama *fundamental*. Desde el punto de vista de la operación de adición de los vectores dispuestos en un eje, la identidad significa que

$$(\overrightarrow{AB}+\overrightarrow{BC})=(\overrightarrow{AB})+(\overrightarrow{BC}). \qquad (9.5)$$

La magnitud de un segmento dirigido determina, salvo el traslado paralelo, el propio segmento en el eje. Mas, al tomar en consideración que los segmentos dirigidos iguales también quedan determinados, salvo el traslado paralelo, quiere decir que la magnitud de un segmento dirigido define unívocamente en el eje dado toda la totalidad de segmentos dirigidos iguales.

Supongamos ahora que se han dado el segmento dirigido $\overrightarrow{AB}$ y el número α. Se denomina *producto $\alpha\cdot\overrightarrow{AB}$ del segmento dirigido $\overrightarrow{AB}$ por un número real* α un segmento dirigido, ubicado en el eje que pasa por los puntos A, B, y cuya magnitud es igual a $\alpha\cdot(\overrightarrow{AB})$. De este modo, por definición

$$(\alpha\cdot\overrightarrow{AB})=\alpha\cdot(\overrightarrow{AB}). \qquad (9.6)$$

Para cualesquiera números α, β y cualesquiera segmentos dirigidos a, b la operación de multiplicación del segmento dirigido por un número posee las siguientes propiedades:

$$1\cdot a=a, \qquad \alpha(\beta a)=(\alpha\beta)a,$$

$$(\alpha+\beta)a=\alpha a+\beta a, \qquad \alpha(a+b)=\alpha a+\alpha b.$$

Las primeras tres propiedades son bastante sencillas. Para demostrarlas basta notar que en los miembros izquierdos y derechos de las igualdades figuran los vectores dispuestos en un eje y hacer uso, luego, de las correlaciones (9.5), (9.6). Demostremos la cuarta propiedad. Supongamos, para simplificar, que $\alpha > 0$. Apliquemos los vectores a, b a un punto común y construyamos sobre ellos un paralelogramo cuya diagonal será igual a $a+b$ (fig. 9.2). Al multiplicar los vectores a, b por α, la diagonal del paralelogramo resul-

tará también multiplicado por α, en virtud de la semejanza de las figuras. Pero esto precisamente significa que $\alpha a + \alpha b = \alpha (a + b)$.

Observemos en conclusión que la introducción de la magnitud de un segmento dirigido puede interpretarse como introducción de cierta "función"

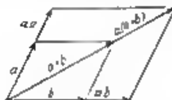


Fig. 9.2.

$$\xi = (x), \tag{9.7}$$

cuyo "argumento" son los vectores x de un mismo eje, mientras que el papel de "valor" lo desempeñan los números reales ξ. En este caso

$$(x + y) = (x) + (y),$$
$$(\lambda x) = \lambda (x)$$

para cualesquiera vectores x e y en el eje y cualquier número λ.

Ejercicios.

1. Demuéstrese que el resultado de la operación de multiplicación por un número no depende de cómo está definida la dirección positiva en el eje.
2. Demuéstrese que el resultado de la operación de multiplicación por un número no depende de cómo está definido el segmento graduado en el eje.
3. Demuéstrese que la operación de multiplicación por un número, definida en cualquier conjunto de segmentos colineales, no hará salir de dicho conjunto.
4. Demuéstrese que la operación de multiplicación por un número, definida en cualquier conjunto de segmentos coplanares, no hará salir de dicho conjunto.
5. ¿Qué representan en sí los segmentos dirigidos opuesto y nulo desde el punto de vista de la operación de multiplicación por un número?

§ 10. Espacios lineales

La resolución de los problemas de cualquier género se reduce, al fin y al cabo, al estudio de ciertos conjuntos y, en primer lugar, al estudio de la estructura de dichos conjuntos. Para estudiar la estructura de conjuntos se emplean los más diversos métodos. Por ejemplo, partiendo de la propiedad característica que poseen los elementos (lo que se hace en los problemas referentes a la construcción de lugares geométricos) o bien partiendo de las propiedades de las operaciones, si están definidas para los elementos.

El último método parece ser de atracción especial debido a su generalidad. Efectivamente, ya hemos visto repetidamente que en los más diversos conjuntos pueden introducirse las más diversas operaciones que, no obstante, poseen propiedades iguales. Será evidente por esta razón que si en la investigación de los conjuntos

se obtiene cierto resultado sólo en la base de las propiedades de una operación, este resultado tendrá lugar en todos los conjuntos, dando las operaciones poseen las mismas propiedades. En este caso la naturaleza concreta tanto de los elementos como de las operaciones sobre ellos puede ser completamente diferente.

Anteriormente hemos introducido para la consideración unos nuevos objetos matemáticos que llamamos segmentos dirigidos o vectores y hemos determinado las operaciones sobre dichos objetos. Se conoce que en realidad detrás de los vectores están los objetos físicos bien reales. Por esto, la investigación detallada de la estructura de los conjuntos representa interés por lo menos para la física.

Ya ahora conocemos tres tipos de conjuntos en los cuales las operaciones poseen las mismas propiedades. Estos conjuntos son: de vectores colineales, de vectores coplanares y de vectores en todo el espacio. A pesar de que en estos conjuntos están introducidas las mismas operaciones, tenemos derecho de esperar que la estructura de los propios conjuntos debe ser diferente.

Hay una tentación, originada por la sencillez de los conjuntos citados, que conduce al deseo de estudiarlos apoyándose sólo en las peculiaridades concretas de los elementos. Sin embargo, no podemos sino observar el hecho de que dichos conjuntos tienen mucho en común, razón por la cual parece racional comenzar su estudio partiendo de ciertas posiciones generales, abrigando esperanza, por lo menos, que *tendremos éxito en evitar repeticiones fastidiosas y monótonas* al pasar de la investigación de un conjunto a la del otro. En adición, esperamos, por supuesto, que en el caso de aparecer algún conjunto con propiedades análogas, a éste último *podríamos extender inmediatamente todos los resultados de las investigaciones realizadas.*

He aquí todos los hechos generales que se conocen acerca de los vectores que forman cualquiera de los tres conjuntos en consideración.

A. A todo par de vectores x, y le corresponde un vector $x + y$, llamado suma de x e y, con la particularidad de que:

1) la adición es conmutativa, $x + y = y + x$,

2) la adición es asociativa, $x + (y + z) = (x + y) + z$,

3) existe un único vector nulo 0 tal que $x + 0 = x$ para cualquier vector x,

4) para todo vector x existe un único vector opuesto $-x$ tal que $x + (-x) = 0$.

B. A todo par α, x (donde α es un número y x, un vector) le corresponde el vector αx, llamado producto de α y x, con la particularidad de que:

1) la multiplicación por los números es asociativa, $\alpha(\beta x) = (\alpha\beta)x$,

2) $1 \cdot x = x$ para todo vector x.

C. Las operaciones de adición y multiplicación están ligadas entre sí por las siguientes correlaciones:

1) la multiplicación por un número es distributiva respecto de la adición de los vectores, $\alpha(x+y) = \alpha x + \alpha y$.

2) la multiplicación por un vector es distributiva respecto de la adición de los números, $(\alpha+\beta)x = \alpha x + \beta x$.

Al igual que en el caso de un campo, los hechos citados no pretenden ser independientes lógicamente. Las propiedades A describen el conjunto de vectores desde el punto de vista de la operación de adición y dicen que respecto de esta operación el conjunto representa un grupo abeliano. Las propiedades B describen el conjunto de vectores desde el punto de vista de la operación de multiplicación de un vector por un número. Las propiedades C describen los lazos existentes entre las dos operaciones.

Examinemos ahora un conjunto K y un campo P de carácter arbitrario. El conjunto K se llamará *espacio lineal* sobre el campo P, si para todos los elementos de K están definidas las operaciones de adición y multiplicación por los números de P y están cumplidos los axiomas A, B, C. De conformidad con esta terminología, podemos decir que el conjunto de vectores colineales, el conjunto de vectores coplanares y el de vectores en todo el espacio son espacios lineales sobre el campo de números reales.

Los elementos de cualquier espacio lineal los llamaremos *vectores*, a pesar de que según su naturaleza concreta dichos elementos pueden ser bien distintos de los segmentos dirigidos. Las representaciones geométricas, relacionadas con el nombre de "vectores", nos ayudarán en ocasiones y, con frecuencia, a prever los resultados necesarios, como también a buscar la interpretación geométrica de diferentes hechos la cual no siempre resulta obvia.

Los vectores del espacio lineal se designarán, como antes, mediante letras latinas minúsculas y los números, con letras minúsculas griegas. Un espacio lineal se llamará *racional*, *real* o *complejo*, según sea el campo P de números racionales, de números reales o de los complejos y se designará mediante D, R y C, respectivamente. El hecho de que en la denominación y la designación no se da ninguna referencia a los elementos del propio espacio tiene un sentido muy profundo de lo cual hablaremos mucho más tarde.

Antes de pasar a la investigación detallada de los espacios lineales, demos a conocer algunos corolarios más sencillos que provienen de la existencia de las operaciones de adición y multiplicación por un número. Conciernen, en lo esencial, a los vectores nulo y opuesto.

En *todo* espacio lineal para cualquier elemento x se verifica la igualdad

$$0\cdot x = 0,$$

donde en el segundo miembro 0 significa el vector nulo y en el primer miembro 0 es el número nulo. Para demostrar esta correlación

examinemos el elemento $0\cdot x + x$. Tenemos

$$0\cdot x + x = 0\cdot x + 1\cdot x = (0 + 1)\,x = 1\cdot x = x.$$

Por consiguiente,

$$x = 0\cdot x + x.$$

Al sumar a ambos miembros de la igualdad el elemento $-x$, hallamos

$$0 = x + (-x) = (0\cdot x + x) + (-x) = \\ = 0\cdot x + (x + (-x)) = 0\cdot x + 0 = 0\cdot x.$$

Ahora ya es fácil indicar la expresión *explícita* para el elemento opuesto $-x$ en términos del propio elemento x. A saber,

$$-x = (-1)\,x.$$

La validez de esta fórmula se deduce de las correlaciones sencillas

$$x + (-1)\,x = 1\cdot x + (-1)\,x = (1 - 1)\,x = 0\cdot x = 0.$$

Esto, a su vez, permite demostrar la validez de las correlaciones

$$-(\alpha x) = (-\alpha)\,x = \alpha\,(-x),$$

puesto que

$$-(\alpha x) = (-1)\,(\alpha x) = (-\alpha)\,x = \alpha\,((-1)\,x) = \alpha\,(-x).$$

Recordemos que por definición de operación de sustracción, $x - y = x + (-y)$ para cualesquiera vectores x e y. La expresión explícita para el vector opuesto permite señalar que las leyes distributivas son legítimas también para *la diferencia*. En efecto, cualesquiera que sean los números α, β y los vectores x, y, tendremos

$$(\alpha - \beta)\,x = \alpha x + (-\beta)\,x = \alpha x + (-(\beta x)) = \alpha x - \beta x,$$

$$\alpha\,(x - y) = \alpha\,(x + (-1)\,y) = \alpha x + (-\alpha)\,y = \\ = \alpha x + (-(\alpha y)) = \alpha x - \alpha y.$$

De aquí proviene, en particular, que para *todo* número

$$\alpha\cdot 0 = 0,$$

puesto que

$$\alpha\cdot 0 = \alpha\,(x - x) = \alpha x - \alpha x = \alpha x + (-\alpha x) = 0.$$

Y, por fin, el último corolario. Si para algún número α y el vector x se verifica la correlación

$$\alpha x = 0, \qquad (10.1)$$

entonces, o *bien* $\alpha = 0$ o *bien* $x = 0$. Efectivamente, si la igualdad (10.1) se realiza, puede haber una de las dos posibilidades: o bien $\alpha = 0$ o bien $\alpha \neq 0$. El caso $\alpha = 0$ confirma nuestra afirmación. Sea,

ahora, $\alpha \neq 0$. En este caso

$$x = 1 \cdot x = \left(\frac{1}{\alpha} \cdot \alpha\right) x = \frac{1}{\alpha}(\alpha x) = \frac{1}{\alpha} \cdot 0 = 0.$$

De aquí se desprende, en particular, que en todo espacio lineal cualquier igualdad *puede ser reducida* formalmente por un factor común no nulo, no importa que dicho factor sea un número o un vector. Efectivamente, si es que $\alpha x = \beta x$ y $x \neq 0$, tenemos $(\alpha - \beta) x = 0$, y entonces $\alpha - \beta = 0$, es decir, $\alpha = \beta$. Si, en cambio, $\alpha x = \alpha y$ y $\alpha \neq 0$, se tiene $\alpha (x - y) = 0$ y entonces $x - y = 0$, es decir, $x = y$.

Así pues, *desde el punto de vista de las operaciones de multiplicación, adición y sustracción tienen lugar formalmente todas las reglas de transformaciones equivalentes para las expresiones algebraicas*. En lo que siguen estas reglas ya no serán objeto de especificaciones especiales.

Así damos por terminado el primer conocimiento con los espacios lineales. Sólo nos resta observar que no es por casualidad que hemos escrito en la forma única las propiedades de operaciones en el campo y en el espacio lineal. Existen rasgos de la semejanza sorprendente (y de la diferencia, en igual medida) entre los axiomas de un campo y de un espacio lineal sobre el campo. El lector ha de reflexionar en esta circunstancia.

Ejercicios.

Demuéstrese que los conjuntos a seguir son espacios lineales. La denominación de la operación refleja siempre no la notación, sino el contenido.

1. El campo: números reales; el conjunto: números reales; la adición: adición de los números reales; la multiplicación por un número: multiplicación de un número real por un número real.

2. El campo: números reales; el conjunto: números complejos; la adición: adición de los números complejos; la multiplicación por un número: multiplicación de un número complejo por un número real.

3. El campo: números racionales; el conjunto: números reales; la adición: adición de los números reales; la multiplicación por un número: multiplicación de un número real por un número racional.

4. El campo: números cualesquiera; el conjunto: un solo vector a; la adición se determina por la regla $a + a = a$; la multiplicación por un número: multiplicación del vector a por un número cualquiera α se determina mediante la regla $\alpha a = a$.

5. El campo: números reales; el conjunto: polinomios con coeficientes reales de una variable t, incluidas las constantes; la adición: adición de los polinomios; la multiplicación por un número: multiplicación del polinomio por un número real.

6. El campo: números racionales, el conjunto: números de la forma $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{5}$, donde a, b, c y d son unos números racionales, la adición: adición de los números de la forma indicada; la multiplicación por un número: multiplicación del número de la forma indicada por un número racional.

7. El campo: campo cualquiera; el conjunto: el mismo campo; la adición: adición de los elementos (¡de los vectores!) del campo; la multiplicación por un número: multiplicación del elemento (¡del vector!) del campo por un elemento (¡un número!) del campo.

§ 11. Sumas finitas y productos finitos

Los campos y los espacios lineales serán conjuntos principales que trataremos en lo sucesivo. En estos conjuntos están introducidas dos operaciones: la adición y la multiplicación. Si se realizan un gran número de operaciones sobre los elementos, aparecen expresiones que contienen en cantidad considerable tanto sumandos como factores. Para asegurar la comodidad de su inscripción introduzcamos símbolos correspondientes. Supondremos, además, que la adición y la multiplicación son operaciones *conmutativas* y *asociativas*.

Sea dado un número finito de los elementos, no necesariamente distintos. Convengamos en considerar que todos los elementos están enumerados de cierta manera y se les han atribuido unos números que varían seguidamente a partir del número k hasta algún número p. Los elementos se designarán mediante una letra, indicando el número. El propio número, el cual se llamará en adelante *índice*, puede ocupar en la designación un lugar arbitrario. El índice puede disponerse al lado de la letra encerrado entre paréntesis, abajo cerca de la letra, arriba cerca de la letra, etc. Esto no tiene ninguna importancia. Con mayor frecuencia lo escribiremos al lado de la letra, abajo, a la derecha.

Una suma de elementos $a_k, a_{k+1}, \ldots, a_p$ se designará mediante el símbolo de la forma siguiente

$$a_k + a_{k+1} + \ldots a_p = \sum_{i=k}^{p} a_i. \tag{11.1}$$

El índice i en esta fórmula se denomina *índice de adición*. Nada variará, por supuesto, si lo designamos con cualquier otra letra. A veces, bajo el signo de la suma se indicará de manera explícita aquella totalidad de los índices según los cuales se realiza la adición. Por ejemplo, la suma en consideración podría ser escrita de la forma:

$$a_k + a_{k+1} + \ldots + a_p = \sum_{k \leqslant i \leqslant p} a_i.$$

Es evidente que si todo elemento a_i es igual al producto del elemento b_i y elemento α, donde α no depende del índice de adición i, entonces

$$\sum_{i=k}^{p} \alpha b_i = \alpha \sum_{i=k}^{p} b_i.$$

es decir, el factor que no depende del índice de adición puede sacarse fuera del signo de la suma.

Supongamos ahora que los elementos están marcados con dos índices cada uno de los cuales varía independientemente. Aceptemos para estos elementos una designación general a_{ij} y sea, por ejemplo, $k \leqslant i \leqslant p$; $m \leqslant j \leqslant n$. Dispongamos los elementos en forma de una tabla rectangular

$$\begin{array}{llll} a_{km} & a_{k,\,m+1} & \ldots & a_{kn} \\ a_{k+1,\,m} & a_{k+1,\,m+1} & \ldots & a_{k+1,\,n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{pm} & a_{p,\,m+1} & \ldots & a_{pn}. \end{array}$$

Está claro que independientemente del orden en que se realiza la sumación, el resultado será el mismo. Por ello, al tomar en consideración la designación introducida para la suma, tenemos

$$(a_{km}+a_{k,\,m+1}+\ldots+a_{kn})+(a_{k+1,\,m}+a_{k+1,\,m+1}+\ldots$$
$$\ldots+a_{k+1,\,n})+\ldots+(a_{pm}+a_{p,\,m+1}+\ldots+a_{pn})=$$
$$=\sum_{j=m}^{n} a_{kj}+\sum_{j=m}^{n} a_{k+1,\,j}+\ldots+\sum_{j=m}^{n} a_{pj}=\sum_{i=k}^{p}\Big(\sum_{j=m}^{n} a_{ij}\Big).$$

Por otro lado, esta misma suma es igual a

$$(a_{km}+a_{k+1,\,m}+\ldots+a_{pm})+(a_{k,\,m+1}+$$
$$+a_{k+1,\,m+1}+\ldots+a_{p,\,m+1})+\ldots+(a_{kn}+a_{k+1,\,n}+\ldots+a_{pn})=$$
$$=\sum_{i=k}^{p} a_{im}+\sum_{i=k}^{p} a_{i,\,m+1}+\ldots+\sum_{i=k}^{p} a_{in}=\sum_{j=m}^{n}\Big(\sum_{i=k}^{p} a_{ij}\Big).$$

Por consiguiente,

$$\sum_{i=k}^{p}\Big(\sum_{j=m}^{n} a_{ij}\Big)=\sum_{j=m}^{n}\Big(\sum_{i=k}^{p} a_{ij}\Big).$$

Si convenimos en realizar la adición siempre de manera sucesiva según los índices de las sumas dispuestas de derecha a la izquierda, los paréntesis pueden ser omitidos y, en definitiva, obtenemos

$$\sum_{i=k}^{p}\sum_{j=m}^{n} a_{ij}=\sum_{j=m}^{n}\sum_{i=k}^{p} a_{ij}.$$

Quiere decir que al sumar según dos índices se puede cambiar el orden de la adición. Si, por ejemplo, $a_{ij}=\alpha_i b_{ij}$, donde α_i no depende del índice j, entonces

$$\sum_{i=k}^{p}\sum_{j=m}^{n} \alpha_i b_{ij}=\sum_{i=k}^{p}\alpha_i\sum_{j=m}^{n} b_{ij}.$$

Los resultados análogos tienen lugar para las sumas realizadas según cualquier número finito de índices.

El producto de los elementos $a_k, a_{k+1}, \ldots, a_p$ se designará mediante el símbolo de tal forma:

$$a_k a_{k+1} \ldots a_p = \prod_{i=k}^{p} a_i.$$

Luego, si $a_i = \alpha b_i$, entonces

$$\prod_{i=k}^{p} \alpha b_i = \alpha^{p-k+1} \prod_{i=k}^{p} b_i.$$

Igual que en el caso de sumación, al calcular un producto según dos índices, se puede cambiar el orden de cálculo del producto, es decir,

$$\prod_{i=k}^{p} \prod_{j=m}^{n} a_{ij} = \prod_{j=m}^{n} \prod_{i=k}^{p} a_{ij}.$$

Todos estos hechos se demuestran según el mismo esquema que se usa en el caso de sumación de los números.

Ejercicios.

Calcúlense las siguientes expresiones:

$$\sum_{i=1}^{n} 1, \quad \sum_{i=k}^{n} 1, \quad \sum_{i=1}^{n} i^2, \quad \sum_{i=1}^{m} i(i-1),$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} ij, \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} (i+3j),$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{p} (2i-j)^2 k,$$

$$\prod_{i=1}^{n} 2, \quad \prod_{j=1}^{n} 10^j, \quad \prod_{i=1}^{n} \prod_{j=1}^{m} 2^{i-j}, \quad \prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{k=1}^{p} 2^{i+j+k}.$$

§ 12. Cálculos aproximados

Los conjuntos examinados son de amplio uso en las más diversas investigaciones teóricas. Para obtener el resultado en dichas investigaciones hemos de realizar casi siempre ciertas operaciones sobre los elementos de los conjuntos. Muy a menudo surge la necesidad en la realización de cálculos con elementos de los campos numéricos. Desearíamos atraer la atención a una *peculiaridad* muy importante *en la realización práctica* de los cálculos semejantes.

Supongamos, para concretar, que tenemos un campo de números reales. Supongamos, además, que todo número se representa como una fracción decimal infinita. Ni un ser humano ni una computadora más moderna pueden operar con las fracciones infinitas. Por esta razón, en la práctica cada fracción de tal índole se sustituye por una fracción decimal finita, próxima a la primera, o por un número racional conveniente.

Así pues, un número real exacto es sustituido por un número aproximado. En las investigaciones teóricas, donde se sobreentiende la prefijación exacta de los números, tal o cual expresión se sustituye bastante a menudo por otra expresión, igual a la primera, aunque escrita, quizás, en una forma distinta. Por supuesto, en este caso tal sustitución no puede despertar objeciones ni tampoco las dudas. En cambio, si deseamos calcular una expresión con ayuda de unos números prefijados aproximadamente, ya no es indiferente para nosotros en qué forma se ha dado la expresión.

Consideremos un ejemplo sencillo. Es fácil comprobar que en el caso de prefijar exactamente el número $\sqrt{2}$ se tiene

$$(\sqrt{2}-1)^6 = 99 - 70\sqrt{2}. \tag{12.1}$$

Puesto que $\sqrt{2} = 1,4142\ldots$, los números $7/5 = 1,4$ y $17/12 = 1,4166\ldots$ pueden considerarse valores aproximados para $\sqrt{2}$. Pero, al sustituir 7/5 en los miembros izquierdo y derecho de (12.1), obtenemos 0,00409 . . . y 1,0, respectivamente. Para 17/12 tenemos 0,00523 . . . y −0,1666 . . . Los resultados de la sustitución son muy distintos uno del otro y no se ve de una tirada cuál de ellos es más próximo al valor exacto. Esto nos muestra *la precaución con la que se deben tratar los números aproximados.*

Nos hemos detenido sólo en una fuente de aparición de los números aproximados que es el redondeo de los números dados con toda la exactitud. En la realidad existen también varias otras fuentes. Por ejemplo, los datos iniciales para los cálculos se obtienen con frecuencia de un experimento, mientras que cualquier experimento puede proporcionar resultados de exactitud limitada. Ya las operaciones más simples, como son la multiplicación y división, pueden conducir al aumento de la cantidad de órdenes en las fracciones. Por ello, nos veremos obligados a despreciar una parte de los órdenes en los resultados de los cálculos intermedios, es decir, aquí también ciertos números se sustituyen por los aproximados, etc.

La investigación detallada de las operaciones con los números aproximados sale de los márgenes de este curso. No obstante, la diferencia entre los cálculos teóricos y prácticos será con frecuencia el objeto de nuestra discusión. La necesidad en tal discusión está provocada por lo que *los cálculos teóricos no pueden ser realizados, como regla, en la forma precisa.*

Ejercicios.

1. ¿Cuál fracción decimal finita es necesaria para aproximar $\sqrt{2}$ para que en los resultados del cálculo de los miembros izquierdo y derecho en (12.1) coincidan los primeros seis órdenes?

2. Supongamos que el resultado de realización de toda operación sobre dos números reales se redondea de acuerdo con cualquier regla conocida hasta t órdenes después de la coma. ¿Quedan en vigor las propiedades de conmutatividad y asociatividad de las operaciones?

3. ¿Se cumplirán en las condiciones del ejercicio antecedente las leyes distributivas?

4. ¿A qué deducción llega Ud., si en los ejercicios 2 y 3 se han obtenido respuestas negativas?

CAPÍTULO 2 ESTRUCTURA DEL ESPACIO LINEAL

§ 13. Combinaciones lineales y cápsulas

Supongamos que en el espacio lineal K, definido sobre un campo P, se ha elegido un número finito de vectores arbitrarios $e_1, e_2, \ldots, e_n$ que no son obligatoriamente distintos. Llamaremos estos vectores *sistema de vectores*. Un sistema de vectores se denominará *subsistema* del segundo sistema, si el primer sistema sólo contiene ciertos vectores del segundo y no contiene ningunos otros vectores.

Sobre los vectores del sistema dado y los vectores obtenidos de los primeros se realizarán las operaciones de adición y multiplicación por unos números. Está claro que todo vector x del tipo

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n, \qquad (13.1)$$

donde $\alpha_1, \alpha_2, \ldots, \alpha_n$ son unos números del campo P, se obtiene de los vectores del sistema dado $e_1, e_2, \ldots, e_n$ con ayuda de las operaciones citadas. Más aún, cualquiera que sea el orden en que se realicen estas operaciones, obtendremos solamente los vectores del tipo (13.1).

Respecto del vector x de (13.1) suele decirse que *se expresa linealmente* en términos de los vectores $e_1, e_2, \ldots, e_n$. El segundo miembro de (13.1) se denomina *combinación lineal* de estos vectores y los números $\alpha_1, \alpha_2, \ldots, \alpha_n$ son *coeficientes de la combinación lineal*.

Fijemos el sistema de vectores $e_1, e_2, \ldots, e_n$ y dejemos que los coeficientes de las combinaciones lineales tomen cualesquiera valores del campo P. En este caso quedará definido cierto conjunto de vectores de K. Este conjunto lleva el nombre de *cápsula lineal* de los vectores $e_1, e_2, \ldots, e_n$ y se designa con $L(e_1, e_2, \ldots, e_n)$.

Nuestro interés hacia las cápsulas lineales se debe a dos circunstancias. En primer lugar, toda cápsula lineal tiene una estructura muy simple: representa en sí una totalidad de todas las combinaciones lineales de vectores del sistema dado. Segundo, la cápsula lineal de cualquier sistema de vectores de todo espacio lineal es de por sí un espacio lineal.

Efectivamente, el cumplimiento de todos los axiomas de un espacio lineal es casi obvio. Sólo los axiomas referentes al vector nulo y al opuesto necesitan, quizás, algunas explicaciones. El vector nulo pertenece a ciencia cierta a cualquier cápsula lineal y corresponde a los valores nulos de los coeficientes de la combinación lineal, es decir,

$$0 = 0\cdot e_1 + 0\cdot e_2 + \ldots + 0\cdot e_n.$$

El vector, opuesto para el vector (13.1), tendrá la forma

$$-x = (-\alpha_1)\, e_1 + (-\alpha_2)\, e_2 + \ldots + (-\alpha_n)\, e_n.$$

La unicidad de los vectores nulo y opuesto proviene de su unicidad como vectores pertenecientes al espacio lineal K.

Observemos que la cápsula lineal de los vectores $e_1, e_2, \ldots, e_n$ es un espacio lineal "*mínimo*" que contiene dichos vectores. En efecto, la propia cápsula lineal sólo consiste de las combinaciones lineales de los vectores $e_1, e_2, \ldots, e_n$, mientras que todo espacio lineal, en el que están contenidos los vectores $e_1, e_2, \ldots, e_n$, ha de contener también todas sus combinaciones lineales.

Así pues, todo espacio lineal contiene, en el caso general, un conjunto infinito de otros espacios lineales representados por las cápsulas lineales. Ahora surgen unas preguntas:

¿En qué condiciones las cápsulas lineales de dos sistemas *diferentes* de vectores consisten de *los mismos* vectores del espacio inicial?

¿Qué número *mínimo* de vectores define una misma cápsula lineal?

¿Será el espacio lineal inicial una cápsula lineal de algunos de sus vectores?

Las respuestas a estas preguntas y otras las daremos en el futuro más próximo. Con este fin emplearemos en gran escala la noción de combinación lineal y, en particular, la propiedad de su *transitividad*. A saber, si un cierto vector z es una combinación lineal de los vectores $x_1, x_2, \ldots, x_r$ y cada uno de ellos, a su turno, es una combinación lineal de los vectores $y_1, y_2, \ldots, y_s$, entonces el vector z también puede ser representado como combinación lineal de los vectores $y_1, y_2, \ldots, y_s$. Demostremos esta propiedad. Sea

$$z = \sum_{i=1}^{r} \beta_i x_i \qquad (13.2)$$

y, además, para todos los números i, $1 \leqslant i \leqslant r$,

$$x_i = \sum_{j=1}^{s} \gamma_{ij} y_j,$$

donde β_i y γ_{ij} son unos números arbitrarios del campo P.

Al sustituir la expresión para x_i en el segundo miembro de (13.2) y al hacer uso de las propiedades correspondientes de las sumas fini-

tes, obtenemos

$$x=\sum_{i=1}^{k}\beta_i x_i=\sum_{i=1}^{k}\beta_i\sum_{j=1}^{m}\gamma_{ij}y_j=\sum_{i=1}^{k}\sum_{j=1}^{m}\beta_i\gamma_{ij}y_j=$$

$$=\sum_{j=1}^{m}\sum_{i=1}^{k}\beta_i\gamma_{ij}y_j=\sum_{j=1}^{m}\Big(\sum_{i=1}^{k}\beta_i\gamma_{ij}\Big)y_j=\sum_{j=1}^{m}\xi_j y_j,$$

donde los coeficientes ξ_j significan las siguientes expresiones:

$$\xi_j=\sum_{i=1}^{k}\beta_i\gamma_{ij}.$$

Por consiguiente, la transitividad de la noción de combinación lineal realmente tiene lugar.

Ejercicios.

1. ¿Qué son en un espacio de segmentos dirigidos las cápsulas lineales de los sistemas de uno, dos, tres y de mayor número de segmentos dirigidos?
2. Examinemos un espacio lineal de polinomios que dependen de la variable t y están definidos sobre el campo de números reales. ¿Qué representa en él la cápsula lineal del sistema de vectores t^2+1, t^2+t, 1?
3. ¿En qué espacio todas las cápsulas lineales coinciden con el mismo espacio?
4. Demuéstrese que un espacio lineal de todos los segmentos dirigidos no puede ser cápsula lineal de dos segmentos dirigidos cualesquiera.

§ 14. Dependencia lineal

Consideraremos otra vez los vectores arbitrarios $x_1, x_2, \dots, x_k$ en un espacio lineal. Puede ocurrir que uno de ellos se exprese linealmente en términos de los demás. Sea, por ejemplo, el vector x_k. Entonces, cada uno de los vectores $x_1, x_2, \dots, x_k$ se expresa linealmente en términos de $x_1, x_2, \dots, x_{k-1}$. Por esta razón cualquier combinación lineal de los vectores $x_1, x_2, \dots, x_k$ es también una combinación lineal de los vectores $x_1, x_2, \dots, x_{k-1}$. Por consiguiente, las cápsulas lineales de los vectores $x_1, x_2, \dots, x_k$ y $x_1, x_2, \dots, x_{k-1}$ coinciden.

Supongamos luego que entre los vectores $x_1, x_2, \dots, x_{k-1}$ hay un vector, por ejemplo, x_{k-1} que también se expresa linealmente en términos de los vectores restantes. Al repetir nuestros razonamientos, llegamos a la conclusión de que ahora cualquier combinación lineal de los vectores $x_1, x_2, \dots, x_k$ es también una combinación lineal de los vectores $x_1, x_2, \dots, x_{k-2}$. Continuando este proceso, pasaremos, al fin y al cabo, del sistema $x_1, x_2, \dots, x_k$ a un sistema de vectores del cual ya *no podremos* excluir ni uno de los vectores. La cápsula lineal del nuevo sistema de vectores *coincide*, evidentemente, con la cápsula lineal de los vectores $x_1, x_2, \dots, x_k$. Además, podemos

decir que si entre $e_1, e_2, \ldots, e_n$ hubo aunque un solo vector no nulo, el nuevo *sistema de vectores o bien consiste solamente en un vector no nulo o bien ninguno de sus vectores se expresa linealmente en términos de los vectores restantes.*

Tal sistema de vectores se llama *linealmente independiente*.

Si un sistema de vectores no es linealmente independiente, se denomina *linealmente dependiente*. En particular, según se infiere de la definición, un sistema compuesto sólo por un vector nulo será linealmente dependiente. La dependencia y la independencia lineales constituyen las propiedades del sistema de vectores. Sin embargo, muy a menudo los adjetivos correspondientes se aplican también a los mismos vectores, razón por la cual en lugar del "sistema linealmente independiente de vectores" diremos, a veces, "sistema de los vectores linealmente independientes", etc.

Desde el punto de vista de las nociones introducidas los razonamientos realizados significan que se ha demostrado el

LEMA 14.1. *Si entre los vectores $e_1, e_2, \ldots, e_n$ no todos son nulos y si dicho sistema es linealmente dependiente, entonces en éste puede existir un subsistema linealmente independiente de vectores, en cuyos términos es linealmente expresado cualquiera de los vectores $e_1, e_2, \ldots, e_n$.*

Una circunstancia, inesperada a primera vista, determina si un sistema de vectores $e_1, e_2, \ldots, e_n$ es linealmente dependiente o linealmente independiente. Ya se ha observado que el vector nulo pertenece a la cápsula lineal y es representado a ciencia cierta por la combinación lineal (13.1) con valores nulos de los coeficientes. A pesar de esto, puede expresarse linealmente en términos de los vectores $e_1, e_2, \ldots, e_n$ de otro método, es decir, determinarse por otro surtido de los coeficientes de la combinación lineal. La independencia lineal de los vectores $e_1, e_2, \ldots, e_n$ está estrechamente vinculada con la unicidad de la representación del elemento nulo en términos de los vectores citados. A saber, es válido el

TEOREMA 14.1. *Un sistema de los vectores $e_1, e_2, \ldots, e_n$ es linealmente independiente cuando, y sólo cuando, de la igualdad*

$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n = 0 \qquad (14.1)$$

se deduce que todos los coeficientes de la combinación lineal son nulos.

DEMOSTRACIÓN. Sea $n = 1$. Si $e_1 \neq 0$, entonces, según se ha mostrado antes, de la correlación $\alpha_1 e_1 = 0$ debe provenir que $\alpha_1 = 0$. En cambio, si de la igualdad $\alpha_1 e_1 = 0$ se desprende que el coeficiente α_1 es nulo, entonces, evidentemente, e_1 no puede ser nulo.

Examinemos ahora el caso en que $n \geqslant 2$. Sea el sistema de vectores linealmente independiente. Supongamos que la igualdad (14.1) resulta lícita con cierto surtido de los coeficientes, entre los cuales existe aunque uno que ser distinto de cero. Sea, por ejemplo, $\alpha_1 \neq 0$.

Entonces, de (14.1) obtenemos

$$e_1 = \left(-\frac{\alpha_2}{\alpha_1}\right)e_2 + \left(-\frac{\alpha_3}{\alpha_1}\right)e_3 + \ldots + \left(-\frac{\alpha_n}{\alpha_1}\right)e_n,$$

es decir, el vector e_1 se expresa linealmente en términos de los demás vectores del sistema. Esto contradice a la condición de independencia lineal del sistema, por lo cual no puede haber coeficientes no nulos entre aquellos que satisfacen (14.1).

Si de la igualdad (14.1) se deduce que todos los coeficientes son nulos, el sistema de vectores no puede ser linealmente dependiente. En efecto, supondremos lo contrario y sea, por ejemplo, un vector e_1 que se expresa linealmente en términos de los demás, es decir,

$$e_1 = \beta_2 e_2 + \beta_3 e_3 + \ldots + \beta_n e_n.$$

En este caso la igualdad (14.1) será satisfecha a ciencia cierta por los coeficientes $\alpha_1 = -1$, $\alpha_2 = \beta_2$, ..., $\alpha_n = \beta_n$, entre los cuales por lo menos uno no es nulo. El teorema queda demostrado.

Este teorema es de tan amplio uso en diversas investigaciones que muy a menudo su afirmación se considera simplemente como definición de independencia lineal.

He aquí dos propiedades sencillas de los sistemas de vectores vinculadas con la dependencia lineal.

LEMA 14.2. *Si algunos de los vectores $e_1, e_2, \ldots, e_n$ son linealmente dependientes, todo el sistema $e_1, e_2, \ldots, e_n$ será linealmente dependiente.*

DEMOSTRACIÓN. Sin restringir la generalidad podemos considerar que los primeros vectores $e_1, e_2, \ldots, e_k$ son linealmente dependientes. Por consiguiente, existen tales números $\alpha_1, \alpha_2, \ldots, \alpha_k$, entre los cuales hay distintos de cero, que

$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_k e_k = 0.$$

De aquí fluye la legitimidad de la igualdad

$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_k e_k + 0 \cdot e_{k+1} + \ldots + 0 \cdot e_n = 0.$$

Mas, esta igualdad significa dependencia lineal de los vectores $e_1, e_2, \ldots, e_n$, puesto que entre los números $\alpha_1, \alpha_2, \ldots, \alpha_k, 0, \ldots, 0$ hay algunos que no son nulos.

LEMA 14.3. *Si entre los vectores $e_1, e_2, \ldots, e_n$ hay aunque un solo vector nulo, todo el sistema $e_1, e_2, \ldots, e_n$ será linealmente dependiente.*

DEMOSTRACIÓN. En efecto, un sistema compuesto por un vector nulo es linealmente dependiente. Por ello, de la propiedad que acabamos de demostrar se infiere dependencia lineal de todo el sistema.

El teorema que sigue representa un resultado más importante referente a la dependencia lineal.

TEOREMA 14.2. *Los vectores $e_1, e_2, \ldots, e_n$ son linealmente dependientes cuando, y sólo cuando, o bien $e_1 = 0$ o bien cierto vector e_k, $2 \leqslant k \leqslant n$, sea una combinación lineal de los vectores antecedentes.*

DEMOSTRACIÓN. Supongamos que los vectores $e_1, e_2, \ldots, e_n$ son linealmente dependientes. Entonces en la igualdad (14.1) no todos los coeficientes son nulos. Sea α_k el último coeficiente no nulo. Si $k = 1$, quiere decir $e_1 = 0$. Sea ahora $k > 1$. En este caso de la igualdad

$$\alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_k e_k = 0$$

encontramos que

$$e_k = \left(-\frac{\alpha_1}{\alpha_k}\right) e_1 + \left(-\frac{\alpha_2}{\alpha_k}\right) e_2 + \ldots + \left(-\frac{\alpha_{k-1}}{\alpha_k}\right) e_{k-1}.$$

Con esto queda demostrada la necesidad de la afirmación enunciada en el teorema. La suficiencia es evidente, puesto que tanto el caso cuando $e_1 = 0$, como el caso en que e_k se expresa linealmente en términos de los vectores anteriores, significa la dependencia lineal de los primeros vectores en el sistema $e_1, e_2, \ldots, e_n$. De aquí se deduce la dependencia lineal de todo el sistema de vectores.

Ejercicios.

1. Demuéstrese que si algún vector de un espacio lineal se representa de modo único en forma de una combinación lineal de los vectores $e_1, e_2, \ldots, e_n$, entonces el sistema citado de vectores es linealmente independiente.
2. Demuéstrese que si el sistema de vectores $e_1, e_2, \ldots, e_n$ es linealmente independiente, entonces cualquier vector de la cápsula lineal de dichos vectores se representa de modo único en forma de su combinación lineal.
3. Demuéstrese que el sistema de vectores $e_1, e_2, \ldots, e_n$ es linealmente dependiente cuando, y sólo cuando, o bien $e_n = 0$, o bien cierto vector e_k, $1 \leqslant k \leqslant n - 1$, es una combinación lineal de los vectores posteriores.
4. Consideraremos un espacio lineal de polinomios que dependen de la variable t y están definidos sobre un campo de números reales. Demuéstrese que el sistema de vectores $1, t, t^2, \ldots, t^n$ es linealmente independiente para todo n.
5. Demuéstrese que un sistema de dos segmentos dirigidos no colineales es linealmente independiente.

§ 15. Sistemas equivalentes de vectores

Examinaremos dos sistemas de vectores de un espacio lineal K. Supongamos que las cápsulas lineales de estos sistemas coinciden y constituyen cierto conjunto L. Al conjunto L le pertenece a sabiendas cierta cualquiera de los vectores de ambos sistemas y, además, cada uno de los vectores de L puede ser representado, en el caso dado, en forma de la combinación lineal tanto de los vectores de un sistema como de los vectores del otro. Por consiguiente:

Dos sistemas de vectores poseen la propiedad de que cualquier vector de cada sistema se expresa linealmente en términos de los vectores del otro sistema.

Los sistemas de tal tipo se llaman *equivalentes.*

De lo dicho se desprende que si las cápsulas lineales de dos sistemas de vectores coinciden, entonces los sistemas son equivalentes. Sean dados, ahora, cualesquiera dos sistemas equivalentes. En virtud de que la noción de combinación lineal es transitiva, toda combinación lineal de los vectores de un sistema puede representarse como combinación lineal de los vectores del otro sistema, es decir, las cápsulas lineales de ambos sistemas coinciden. Así pues, es válido el

LEMA 15.1. *Para que las cápsulas lineales de dos sistemas de vectores coincidan, es necesario y suficiente que dichos sistemas sean equivalentes.*

Hemos de notar que el concepto de equivalencia de dos sistemas de vectores es *una relación de equivalencia*. La reflexividad es obvia, puesto que todo sistema es equivalente a sí mismo; la simetría fluye de la definición de sistemas equivalentes y la transitividad de este concepto proviene de la transitividad de la noción de combinación lineal. Por esta razón el *conjunto de todos los sistemas de vectores de un espacio lineal se puede dividir en clases de los sistemas equivalentes.* Es importante subrayar que a todos los sistemas de cada clase les corresponde *una y sólo una* cápsula lineal.

Sobre la proporción del número de vectores en los sistemas equivalentes nada puede decirse en el caso general. En cambio, si de los dos sistemas equivalentes aunque uno es linealmente independiente, podemos hacer deducciones definitivas acerca de la cantidad de los vectores. Estas deducciones se fundamentan en el

TEOREMA 15.1. *Si cada uno de los vectores de un sistema linealmente independiente* $e_1, e_2, \ldots, e_n$ *se expresa linealmente en términos de los vectores* $y_1, y_2, \ldots, y_m$, *entonces* $n \leq m$.

DEMOSTRACIÓN. Por hipótesis del teorema, el vector e_n se expresa linealmente en términos de los vectores $y_1, y_2, \ldots, y_m$, por lo cual el sistema

$$e_n, y_1, y_2, \ldots, y_m \tag{15.1}$$

es linealmente dependiente. El vector e_n no es nulo, razón por la cual, de acuerdo con el teorema 14.2, cierto vector y_i de (15.1) es una combinación lineal de los vectores anteriores. Al eliminar este vector, obtendremos el sistema:

$$e_n, y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m. \tag{15.2}$$

Ahora, haciendo uso de la transitividad de la noción de combinación lineal, es fácil mostrar que cada uno de los vectores $e_1, e_2, \ldots, e_n$ se expresa linealmente en términos de los vectores (15.2).

Adjuntamos a los vectores (15.2) por la izquierda el vector e_{k-1}. De nuevo llegamos a que el sistema

$$e_{k-1},\ e_k,\ y_1,\ \ldots,\ y_{l-1},\ y_{l+1},\ \ldots,\ y_m \qquad (15.3)$$

es linealmente dependiente. El vector e_{k-1} no es nulo, por lo cual, de acuerdo con el teorema 14.2, uno de los vectores restantes (15.3) es una combinación lineal de los vectores anteriores. De este vector no puede servir e_k, puesto que en tal caso sería linealmente dependiente el sistema de dos vectores e_{k-1}, e_k y, por consiguiente, todo el sistema $e_1, e_2, \ldots, e_k$. De este modo, cierto vector y_j de (15.3) se expresa linealmente en términos de los anteriores. Si lo eliminamos, obtendremos de nuevo un sistema por cuyo intermedio se expresa linealmente cada uno de los vectores $e_1, e_2, \ldots, e_k$.

Continuando este proceso advertimos que los vectores $y_1, y_2, \ldots, y_m$ no pueden ser agotados antes de que se hayan adjuntado todos los vectores $e_1, e_2, \ldots, e_k$. En el caso contrario resulta que cada uno de los vectores $e_1, e_2, \ldots, e_k$ se expresa linealmente en términos de una parte de los vectores de este mismo sistema, es decir, todo el sistema ha de ser linealmente dependiente. Puesto que esto contradice a la hipótesis del teorema, de aquí se deduce que $n \leqslant m$.

Consideraremos unos corolarios de este teorema. Sean dados dos sistemas equivalentes de vectores linealmente independientes. De acuerdo con el teorema demostrado, cada uno de dichos sistemas contiene no más vectores que el otro. Por consiguiente:

Los sistemas equivalentes linealmente independientes se componen de un mismo número de vectores.

Luego, tomemos n vectores arbitrarios, construyamos en ellos una cápsula lineal y elijamos en ésta $n+1$ vectores cualesquiera. Como el número de estos vectores es mayor que el de vectores dados, ellos no pueden ser linealmente independientes. Por esto:

Cualesquiera $n+1$ vectores de la cápsula lineal de un sistema de n vectores son linealmente dependientes.

La afirmación del lema 14.1 significa en términos de los sistemas equivalentes que cualquiera que sea el sistema de los vectores que no son nulos simultáneamente, existe en él un subsistema linealmente independiente que es equivalente al sistema. Este subsistema lleva el nombre de *base* del sistema inicial.

Por supuesto, todo sistema puede tener más que una base. Todas las bases de los sistemas equivalentes son de por sí sistemas equivalentes. Del primer corolario del teorema 15.1 se infiere que constan de un mismo número de vectores. Este número es la característica de todos los sistemas equivalentes y se llama *rango*. Por definición, el rango de los sistemas de vectores nulos se considera igual a cero.

Examinemos ahora dos sistemas linealmente independientes que se componen de un mismo número de vectores. Sustituyamos un

vector cualquiera del primer sistema por cierto vector del segundo sistema. A continuación, en el sistema obtenido sustituyamos otra vez uno de los vectores del primer sistema por alguno de los vectores restantes del segundo sistema, etc. El proceso de sustituciones se realizará hasta que el primer sistema sea sustituido por el segundo sistema. Si la sustitución se realiza de un modo arbitrario, los sistemas intermedios pueden resultar linealmente dependientes. No obstante, queda válido el

TEOREMA 15.2. *El proceso de la sustitución sucesiva puede realizarse de una manera tal que todos los sistemas intermedios serán linealmente independientes.*

DEMOSTRACIÓN. Sean dados dos sistemas linealmente independientes de los vectores $y_1, y_2, \ldots, y_n$ y $z_1, z_2, \ldots, z_n$. Supongamos que se han realizado k pasos del proceso, donde $k \geqslant 0$. Sin restringir la generalidad consideraremos que todos los vectores $y_1, \ldots, y_k$ están sustituidos por los vectores $z_1, \ldots, z_k$ y todos los sistemas obtenidos, incluido el sistema

$$z_1, \ldots, z_k, y_{k+1}, \ldots, y_n,$$

son linealmente independientes. Esta suposición tiene lugar a ciencia cierta cuando $k = 0$.

Supongamos luego que durante la sustitución del vector y_{k+1} por cualquiera de los vectores $z_{k+1}, \ldots, z_n$ todos los sistemas

$$z_1, \ldots, z_k, z_i, y_{k+2}, \ldots, y_n$$

son linealmente dependientes para $i = k + 1, \ldots, n$. Dado que el sistema

$$z_1, \ldots, z_k, y_{k+2}, \ldots, y_n \qquad (15.4)$$

es linealmente independiente, resulta que los vectores z_i, para $i = k + 1, \ldots, n$, se expresan de modo lineal en términos de este sistema. Pero, a través del mismo se expresan linealmente, además, los vectores z_i cuando $i = 1, 2, \ldots, k$. Por consiguiente, a través del sistema (15.4) deben expresarse linealmente todos los vectores $z_1, \ldots, z_n$. Esto no es posible en virtud del teorema 15.1. Por esta razón, el proceso de sustitución enunciado en el teorema realmente tiene lugar.

Ejercicios.

Demuéstrese que las transformaciones del sistema de vectores a seguir, que se llaman *elementales*, conducen a un sistema equivalente.

1. La adjunción al sistema de vectores de cualquier combinación lineal de estos vectores.
2. La eliminación del sistema de vectores de cualquier vector que es una combinación lineal de los vectores restantes.
3. La multiplicación de cualquier vector de un sistema por un número distinto de cero.
4. La adición a cualquier vector de un sistema de cualquier combinación lineal de los vectores restantes.
5. La permutación de dos vectores.

§ 16. Base

Sea dado un espacio lineal arbitrario que se compone *no sólo de un vector nulo*. En tal espacio se tiene a ciencia cierta aunque un vector no nulo y, por lo tanto, existe un sistema linealmente independiente, por lo menos, de un vector. Por consiguiente, son posibles dos casos: o bien existe un sistema linealmente independiente que contiene un número de vectores tan grande como se quiera o bien existe un sistema linealmente independiente que contiene el número máximo de vectores. En el primer caso el espacio lineal se llama *de dimensión infinita* y en el segundo caso, *de dimensión finita*.

A excepción de unos ejemplos episódicos, *nuestra atención será dirigida a lo largo de todo el curso exclusivamente a los espacios de dimensión finita*. En particular, un espacio lineal de dimensión finita lo constituirá cualquier cápsula lineal construida en el número finito de vectores de un espacio arbitrario (no necesariamente de dimensión finita).

Así pues, supongamos que en el espacio lineal de dimensión finita K los vectores $e_1, e_2, \ldots, e_n$ constituyen un sistema linealmente independiente con el número máximo de vectores. Esto significa que para cualquier vector x de K el sistema $e_1, e_2, \ldots, e_n, x$ será linealmente dependiente. De conformidad con el teorema 14.2, el vector x se expresa linealmente en términos de $e_1, e_2, \ldots, e_n$. Como el vector x es arbitrario, mientras que los vectores $e_1, e_2, \ldots, e_n$ son fijados, podemos decir que:

Cualquier espacio lineal de dimensión finita es una cápsula lineal del número finito de sus vectores.

Ahora al investigar espacios lineales de dimensión finita, podemos hacer uso de toda clase de información referente a las cápsulas lineales y los sistemas equivalentes de vectores. Introduzcamos la siguiente definición.

Un sistema linealmente independiente de los vectores en términos de los cuales se expresa linealmente todo vector del espacio, se llama *base del espacio*.

La noción de base está ligada aquí con un sistema linealmente independiente que contiene el número máximo de vectores. No obstante, es evidente que todas las bases de un mismo espacio lineal de dimensión finita representan sistemas equivalentes linealmente independientes. Como sabemos, tales sistemas contienen un número igual de vectores. Por consiguiente, el número de los vectores de una base es la característica del espacio lineal de dimensión finita. Este número recibe el nombre de *dimensión* del espacio lineal K y se designa $\dim K$. Si $\dim K = n$, el propio espacio K se llama *n*-dimensional. Está claro que:

En un espacio lineal n-dimensional todo sistema linealmente inde-

pendiente de n vectores forma una base, mientras que todo sistema de $n+1$ vectores es linealmente dependiente.

Observemos que en todos los razonamientos anteriores suponíamos que el espacio lineal consiste no sólo en un vector nulo. Un espacio, que contiene sólo un vector nulo, no posee base en nuestro sentido y, rigiéndonos por la definición, consideraremos que su dimensión es igual a cero.

La base es de importancia enorme en el estudio de los espacios lineales de dimensión finita y en nuestras investigaciones siempre vamos a utilizarla. La base permite describir con facilidad la estructura de cualquier espacio lineal definido sobre un campo arbitrario P. Además, con ayuda de la base se puede construir un *aparato eficaz que reduce la realización de las operaciones sobre los elementos del espacio a las operaciones correspondientes sobre los números del campo P.*

Como ya hemos mostrado más arriba, todo vector x del espacio lineal K puede ser representado en forma de una combinación lineal

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n, \tag{16.1}$$

donde, $\alpha_1, \alpha_2, \ldots, \alpha_n$ son ciertos números de P y $e_1, e_2, \ldots, e_n$ la base de K. La combinación lineal (16.1) se denomina *descomposición del vector x según la base* y los propios números $\alpha_1, \alpha_2, \ldots, \alpha_n$ se llaman *coordenadas del vector x respecto de dicha base.* El hecho de que el vector x se ha dado mediante sus coordenadas $\alpha_1, \alpha_2, \ldots, \alpha_n$ se notará del modo siguiente:

$$x = (\alpha_1, \alpha_2, \ldots, \alpha_n).$$

En lo que sigue no se indica, por regla general, a qué base se refieren las coordenadas dadas, siempre que esto no lleve a la ambigüedad.

Es fácil mostrar que para todo vector x de K *su descomposición según la base es única.* Esto se demuestra por un procedimiento empleado muy a menudo en la resolución de los problemas referentes a la dependencia lineal. Supongamos que existe otra descomposición:

$$x = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n. \tag{16.2}$$

Al restar término a término (16.2) de (16.1), obtenemos

$$(\alpha_1 - \beta_1) e_1 + (\alpha_2 - \beta_2) e_2 + \ldots + (\alpha_n - \beta_n) e_n = 0.$$

En virtud de que los vectores $e_1, e_2, \ldots, e_n$ son linealmente independientes, de aquí se deduce que todos los coeficientes de la combinación lineal son nulos y que, por consiguiente, las descomposiciones (16.1) y (16.2) coinciden.

De este modo, con la base fijada del espacio lineal K todo vector de K se define unívocamente por la totalidad de sus coordenadas respecto de esta base.

Supongamos ahora que cualesquiera dos vectores x e y de K vienen dados por sus coordenadas respecto de una misma base e_1,

$e_1, \ldots, e_n$, es decir,

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$
$$y = \gamma_1 e_1 + \gamma_2 e_2 + \ldots + \gamma_n e_n,$$

entonces

$$x + y = (\alpha_1 + \gamma_1) e_1 + (\alpha_2 + \gamma_2) e_2 + \ldots + (\alpha_n + \gamma_n) e_n.$$

Luego, para cualquier número λ del campo P tenemos

$$\lambda x = (\lambda\alpha_1) e_1 + (\lambda\alpha_2) e_2 + \ldots + (\lambda\alpha_n) e_n.$$

De aquí proviene que *al adicionar dos vectores de un espacio lineal, sus coordenadas respecto de cualquier base se suman y al multiplicar un vector por un número, todas sus coordenadas se multiplican por dicho número.*

Ejercicios.

1. Demuéstrese que el rango de un sistema de vectores coincide con la dimensión de su cápsula lineal.
2. Demuéstrese que los sistemas equivalentes de vectores tienen un mismo rango.
3. Demuéstrese que si la cápsula lineal L_1 está construida en los vectores de la cápsula lineal L_2, entonces $\dim L_1 \leqslant \dim L_2$.
4. Demuéstrese que si la cápsula lineal L_1 está construida en los vectores de la cápsula lineal L_2 y si $\dim L_1 = \dim L_2$, entonces las propias cápsulas lineales coinciden.
5. Demuéstrese que un espacio lineal de polinomios con coeficientes reales dado sobre un campo de números reales es de dimensión finita.

§ 17. *Ejemplos sencillos de los espacios lineales*

Los conceptos fundamentales de dependencia lineal y de base se pueden ilustrar con unos ejemplos muy sencillos pero aleccionadores, si tomamos a título de espacios lineales los conjuntos de números con operaciones ordinarias de adición y multiplicación. La validez de los axiomas del espacio lineal para los conjuntos de este tipo es completamente obvia, por lo cual no nos detendremos en su comprobación. Los elementos del espacio se llamarán, como antes, *vectores*.

Consideraremos un espacio lineal *complejo* que representa en sí un grupo de adición de todos los números complejos con multiplicación sobre un campo de números complejos. Está claro que cualquier número z_1, distinto de cero, representa en él un vector linealmente independiente. Sin embargo, cualesquiera dos vectores no nulos z_1 y z_2 ya son siempre linealmente dependientes. Para demostrarlo basta hallar tales dos números complejos α_1 y α_2, distintos de cero simultáneamente, que sea $\alpha_1 z_1 + \alpha_2 z_2 = 0$. Mas, esta igualdad puede realizarse, evidentemente, suponiendo $\alpha_1 = -z_2$, $\alpha_2 = z_1$. Por consiguiente, el espacio lineal considerado es unidimensional.

Resulta ser algo diferente un espacio lineal *real* que representa

un grupo de adición de todos los números complejos con multiplicación sobre un campo de números reales. A título de coeficientes de las combinaciones lineales pueden emplearse, en este caso, solamente los números reales, razón por la cual este espacio lineal no puede ser unidimensional. En efecto, no existen números reales α_1 y α_2, distintos de cero simultáneamente, para los cuales la combinación lineal $\alpha_1 z_1 + \alpha_2 z_2$ se reduciría a cero, por ejemplo, para $z_1 = 1$, $z_2 = i$. Sea al cargo del lector demostrar, a título de ejercicio, que el espacio lineal dado es bidimensional.

Es importante recalcar que aunque ambos espacios examinados se componen de *los mismos* elementos, *se diferencian* uno del otro en principio. Ahora ya está claro que un espacio lineal real, que representa un grupo de adición de todos los números reales con multiplicación sobre un campo de números reales, es unidimensional. Consideremos luego un espacio lineal racional que representa en sí un grupo de adición de todos los números reales con multiplicación sobre un campo de números racionales.

Tratemos de construir, como antes, un sistema que contenga el número máximo de los vectores linealmente independientes x_1, x_2, x_3, ... Está claro que se puede tomar, por ejemplo, $x_1 = 1$. Puesto que a título de coeficientes de las combinaciones lineales α_1, α_2, α_3, ... se admiten sólo los números racionales, es natural que mediante un número del tipo $\alpha_1 \cdot 1$ no es posible representar, por ejemplo, $\sqrt{2}$. Por lo tanto, el espacio no puede ser unidimensional. Por eso, en calidad del segundo vector, linealmente independiente con la unidad, podremos tomar precisamente $\sqrt{2}$. Sin embargo, mediante un número del tipo $\alpha_1 \cdot 1 + \alpha_2 \sqrt{2}$ no se puede representar, por ejemplo, $\sqrt[4]{2}$.

Efectivamente, supongamos que para ciertos α_1 y α_2 racionales se verifica la igualdad $\sqrt[4]{2} = \alpha_1 + \alpha_2 \sqrt{2}$. Elevando al cuadrado ambos miembros, obtendremos

$$\sqrt{2} = (\alpha_1^2 + 2\alpha_2^2) + 2\alpha_1\alpha_2 \sqrt{2}$$

o bien

$$\frac{\alpha_1^2 + 2\alpha_2^2}{1 - 2\alpha_1\alpha_2} = \sqrt{2}.$$

Esto es imposible, puesto que en el primer miembro figura un número racional y en el segundo, irracional.

Por lo tanto, el espacio en consideración tampoco puede ser bidimensional. Entonces, ¿cuál será? Por sorprendente que sea, es *de dimensión infinita*. Sin embargo, la demostración de esta afirmación sale de los márgenes de nuestro curso.

La atención especial que se ha prestado a los ejemplos de espacios lineales de dimensión pequeña se explica por el hecho de que con ayuda de tales espacios podemos construir los espacios lineales de cualquier dimensión. Mas, hablaremos de esto más tarde.

Ejercicios.

1. ¿Qué dimensión tiene un espacio lineal de números racionales dado sobre un campo de números racionales?
2. Constrúyanse los sistemas linealmente independientes de vectores en un espacio de los números complejos dado sobre un campo de números racionales.
3. ¿Será espacio lineal un grupo de adición de los números racionales dado sobre un campo de números reales? Si no ¿por qué?

§ 18. Espacios lineales de segmentos dirigidos

Ya se ha señalado anteriormente que los conjuntos de segmentos dirigidos colineales, de segmentos dirigidos coplanares y de segmentos dirigidos en todo el espacio forman unos espacios lineales sobre un campo de números reales. Nuestra tarea inmediata consiste en revelar su dimensión y construir la base.

LEMA 18.1. *La condición necesaria y suficiente de la dependencia lineal de dos vectores es su carácter colineal.*

DEMOSTRACIÓN. Hemos de notar que la afirmación del lema es evidente, si entre dos vectores se tiene aunque uno nulo. Por eso supondremos que ambos vectores no son nulos.

Sean a y b dos vectores linealmente dependientes. En este caso existen los números α y β tales que

$$\alpha a + \beta b = 0.$$

Puesto que, de acuerdo con la suposición, $a \neq 0$, $b \neq 0$, entonces $\alpha \neq 0$, $\beta \neq 0$, y por eso

$$b = \left(-\frac{\alpha}{\beta}\right) a.$$

Por consiguiente, según la definición de operación de la multiplicación de un segmento dirigido por un número, los vectores a y b son colineales.

Supongamos ahora que los vectores a y b son colineales. Apliquémoslos a un punto común O. Estos vectores se dispondrán en cierta recta la cual transformaremos en un eje con una dirección determinada. Los vectores a y b son no nulos, razón por la cual existe un número real λ tal que la magnitud del segmento dirigido a es igual al producto de la magnitud del segmento dirigido b por el número λ, es decir, $\{a\} = \lambda\{b\}$. Mas, de acuerdo con la definición de operación de multiplicación de un segmento dirigido por un número, esto quiere decir que $a = \lambda b$. Así pues, los vectores a y b son linealmente dependientes.

Del lema demostrado se desprende que el espacio lineal de segmentos dirigidos colineales es *unidimensional* y de su base puede servir *cualquier vector no nulo.*

El lema 18.1. permite deducir un corolario muy importante. A saber, si los vectores a, b son colineales y $a \neq 0$, entonces existe tal número λ que $b = \lambda a$. En efecto, estos vectores son linealmente dependientes, es decir, para ciertos números α, β que no son nulos simultáneamente, se tiene $\alpha a + \beta b = 0$. Si suponemos que $\beta = 0$, de aquí se deduce que $\alpha = 0$. Por consiguiente, $\beta \neq 0$ y a título de número λ se puede tomar $\lambda = (-\alpha)/\beta$.



Fig. 18.1.

LEMA 18.2. *La condición necesaria y suficiente de la dependencia lineal de tres vectores es su carácter coplanar.*

DEMOSTRACIÓN. Sin restringir la generalidad supondremos que ningún par de los tres vectores indicados es colineal, puesto que en el caso contrario la afirmación del lema se deduce inmediatamente del lema 18.1.

Así pues, sean a, b, c tres vectores linealmente dependientes. Por consiguiente, existen tales números reales α, β, γ, entre los cuales no todos son nulos, que

$$\alpha a + \beta b + \gamma c = 0.$$

Si, por ejemplo, $\gamma \neq 0$, entonces de esta igualdad hallamos

$$c = \left(-\frac{\alpha}{\gamma}\right) a + \left(-\frac{\beta}{\gamma}\right) b.$$

Apliquemos los vectores a, b, c a un punto común O. En este caso de la última igualdad se desprende que el vector c es igual a la diagonal del paralelogramo construido en los vectores $(-\alpha/\gamma)\, a$ y $(-\beta/\gamma)\, b$. Esto significa que siendo trasladados paralelamente en el punto común, los vectores a, b, c resultan ser dispuestos en un mismo plano y, por consiguiente, son coplanares.

Supongamos ahora que los vectores a, b, c son coplanares. Trasladémoslos en un plano y apliquemos al punto común O (fig. 18.1). Por los extremos del vector c tracemos las rectas paralelas a los vectores a y b, y consideremos el paralelogramo $OACB$. Los vectores a, $\overrightarrow{OA}$ y b, $\overrightarrow{OB}$ son colineales por construcción y no nulos, por lo cual existen tales números λ, μ que

$$\overrightarrow{OA} = \lambda a, \quad \overrightarrow{OB} = \mu b.$$

Pero, $\overrightarrow{OC} = \overrightarrow{OA} + \overrightarrow{OB}$, por consiguiente $c = \lambda a + \mu b$, o bien

$$\lambda a + \mu b + (-1)\, c = 0.$$

Puesto que los números λ, μ, -1 son a ciencia cierta diferentes de cero, la última igualdad significa dependencia lineal de los vectores a, b, c.

Ahora podemos resolver la cuestión sobre la dimensión de un espacio lineal de segmentos dirigidos coplanares. De acuerdo con el lema que acabamos de demostrar, la dimensión de este espacio debe ser menor de tres. Pero cualesquiera dos segmentos dirigidos no colineales de dicho espacio son linealmente independientes. Por eso, el espacio lineal de segmentos dirigidos coplanares es un espacio *bidimensional* y su base pueden constituir *cualesquiera dos vectores no colineales*.

LEMA 18.3. *Cualesquiera cuatro vectores son linealmente dependientes.*

DEMOSTRACIÓN. Sin restringir la generalidad supondremos que ninguna terna de cuatro vectores es coplanar, puesto que en el caso contrario la afirmación del lema fluye inmediatamente del lema 18.2. Apliquemos los vectores a, b, c, d al origen común O y por el extremo D del vector d tracemos los planos paralelos a los que se definen por los pares de vectores b, c; a, c y a, b, respectivamente (fig. 18.2). De la regla del paralelogramo para la adición de vectores proviene que

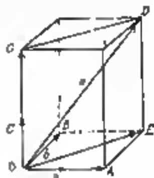


Fig. 18.2.

$$\overrightarrow{OD}=\overrightarrow{OC}+\overrightarrow{OE}, \quad \overrightarrow{OE}=\overrightarrow{OA}+\overrightarrow{OB},$$

por ello

$$\overrightarrow{OD}=\overrightarrow{OA}+\overrightarrow{OB}+\overrightarrow{OC}. \qquad (18.1)$$

Los vectores a, $\overrightarrow{OA}$ y también b, $\overrightarrow{OB}$ y c, $\overrightarrow{OC}$ son colineales según la construcción, con la particularidad de que a, b, c son no nulos. Por consiguiente, existen tales números λ, μ, ν que

$$\overrightarrow{OA}=\lambda a, \quad \overrightarrow{OB}=\mu b, \quad \overrightarrow{OC}=\nu c.$$

Teniendo presente (18.1), esto nos da la correlación

$$d=\lambda a+\mu b+\nu c,$$

de donde se desprende la dependencia lineal de los vectores a, b, c, d.

Del lema demostrado concluimos que la dimensión de un espacio lineal de todos los segmentos dirigidos debe ser inferior a cuatro. Pero no puede ser menos que tres, puesto que, de conformidad con el lema 18.2, cualesquiera tres segmentos dirigidos no coplanares son linealmente independientes. Por eso el espacio lineal de todos los segmentos dirigidos es *tridimensional* y de su base pueden servir *cualesquiera tres vectores no coplanares*.

Los espacios lineales considerados no son muy ilustrativos desde el punto de vista geométrico, puesto que se admite la existencia en dichos espacios de un número infinito de los vectores iguales entre sí. Se harán mucho más ilustrativos, si en toda clase de los vectores iguales fijamos un representante y por palabra "vector" entendemos siempre un segmento dirigido elegido sólo de la totalidad de estos representantes.

Uno de los métodos más cómodos para la fijación consiste en el examen de los conjuntos de vectores dirigidos sujetados en cierto punto O. En tal caso en lugar de un espacio lineal de segmentos dirigidos colineales obtenemos un espacio de segmentos dirigidos que están sujetados en el punto O y se ubican en la recta que pasa por este punto; en lugar de un espacio lineal de segmentos dirigidos coplanares obtenemos un espacio de segmentos dirigidos sujetados al punto O y ubicados en el plano que pasa por este punto; y, por fin, en lugar de un espacio lineal de todos los segmentos dirigidos, un espacio de segmentos dirigidos sujetados en el punto O.

En adelante trataremos, principalmente, solamente los vectores sujetados. Los espacios lineales correspondientes se designarán mediante V_1, V_2 y V_3, donde el índice abajo significa la dimensión. Un espacio lineal compuesto de un solo segmento dirigido nulo, lo designaremos mediante V_0.

La introducción de estos espacios permite establecer una correspondencia biunívoca entre los puntos y los segmentos dirigidos. Para ello es suficiente a todo vector ponerle en correspondencia su punto final. Teniendo presente tal *interpretación geométrica*, los elementos de un espacio lineal abstracto también se llamarán, a veces, *puntos* en lugar de denominarlos vectores.

Ejercicios.

Indíquese en los espacios V_1, V_2, V_3 el sentido geométrico de tales conceptos como:

1. La cápsula lineal.
2. La dependencia e independencia lineales.
3. Los sistemas equivalentes de vectores.
4. Las transformaciones equivalentes elementales del sistema de vectores.
5. El rango del sistema de vectores.

§ 10. Suma e intersección de los subespacios

La introducción de las cápsulas lineales ha mostrado que todo espacio lineal contiene un conjunto infinito de otros espacios lineales. El significado de estos espacios no se limita sólo a las cuestiones examinadas anteriormente.

Las cápsulas lineales se han definido con ayuda de la indicación inmediata de su *estructura*. Podríamos emplear también otro acceso,

al definir los espacios "menores" a través de *las propiedades* de los vectores. Supongamos que en el espacio lineal K está definido un conjunto de vectores L. Si, al realizar las mismas operaciones que en el espacio K, el conjunto L es por sí mismo un espacio lineal, llamaremos a L *subespacio* lineal, subrayando en la denominación el hecho de que el subespacio consta de los vectores de cierto espacio. Por lo visto, el subespacio mínimo es aquel que se compone sólo de un vector nulo. Tal subespacio se denominará *nulo* y se designará mediante el símbolo 0. El subespacio máximo es el espacio K. Estos dos subespacios se llaman *triviales* y los demás, *no triviales*. Es evidente que todo subespacio junto con cada par de sus elementos x, y contiene también todas sus combinaciones lineales $\alpha x + \beta y$. Lo recíproco es también cierto. A saber:

Si el conjunto de vectores L de un espacio lineal K, junto con cada par de sus elementos x, y contiene también *todas* sus combinaciones lineales $\alpha x + \beta y$, será un subespacio.

Efectivamente, de todos los axiomas para el espacio lineal es preciso comprobar sólo aquellos que se refieren a los vectores nulo y opuesto. La validez de los axiomas restantes es obvia. Tomemos $\alpha = 0$, $\beta = 0$. De conformidad con los corolarios que provienen de las propiedades de las operaciones para los vectores del espacio K, llegamos a que $0 \cdot x + 0 \cdot y = 0$, es decir, el vector nulo pertenece al conjunto L. Tomemos ahora $\alpha = -1$, $\beta = 0$. Se tiene $(-1)\,x + 0 \cdot y = (-1)\,x$, por lo cual en L junto con todo vector x figura también un vector opuesto a x. Así pues, el conjunto L es un subespacio.

La presencia de la base permite enunciar la afirmación de que en todo espacio de dimensión finita cualquier subespacio será una cápsula lineal. Por esto, en los espacios lineales de dimensión finita la cápsula lineal constituye un método más general para definir los subespacios lineales. En el espacio de dimensión infinita no es así. No obstante, no se debe olvidar que existe muchísimo en común entre los conceptos y los hechos en los espacios de dimensión finita y los análogos correspondientes en los espacios de dimensión infinita. En nuestro deseo de subrayar esta circunstancia, aun en los espacios de dimensión finita utilizaremos con mayor frecuencia el término *subespacio lineal* en vez de *cápsula lineal*.

Sea K un espacio n-dimensional. Al igual que en el mismo espacio K, en cualquier subespacio suyo L puede construirse la base. Si en el espacio K la base elegida es $e_1, e_2, \ldots, e_n$, en el caso general los vectores básicos del subespacio L no pueden elegirse directamente del número de los vectores $e_1, e_2, \ldots, e_n$ aunque sea por aquella razón que ni uno de ellos pueda figurar en dicho subespacio. Sin embargo, resulta válido en cierto sentido el siguiente lema inverso.

LEMA 10.1. *Si en un subespacio L de dimensión s se ha elegido una base arbitraria $t_1, \ldots, t_s$, entonces en el espacio K de n-ésima*

dimensión se pueden elegir los vectores $t_{r+1}, \ldots, t_n$ de una manera tal que el sistema de los vectores $t_1, \ldots, t_r, t_{r+1}, \ldots, t_n$ será la base en todo K.

DEMOSTRACIÓN. Examinemos sólo aquellos sistemas linealmente independientes de vectores en K que contienen los vectores $t_1, \ldots, t_r$. Está claro que entre estos sistemas hay un sistema $t_1, \ldots, t_r, t_{r+1}, \ldots, t_p$ que tiene el número máximo de vectores. Mas, en este caso cualquiera que sea el vector x de K, el sistema $t_1, \ldots, t_p, x$ debe ser linealmente dependiente. Por consiguiente, el vector x debe expresarse linealmente en términos de los vectores $t_1, \ldots, t_p$. Esto significa que los vectores $t_1, \ldots, t_r, t_{r+1}, \ldots, t_p$ forman la base en K y $p = n$.

Consideremos otra vez el espacio lineal arbitrario K. Este espacio engendra el conjunto de todos los subespacios suyos, el cual se designará mediante U. En el conjunto U se pueden definir dos operaciones algebraicas que, a base de unos subespacios, permiten construir otros.

Se denomina *suma* $L_1 + L_2$ de los subespacios lineales L_1, L_2 un conjunto de todos los vectores del tipo $z = x + y$, donde $x \in L_1$, $y \in L_2$.

Se denomina *intersección* $L_1 \cap L_2$ de los subespacios lineales L_1, L_2 un conjunto de todos los vectores pertenecientes simultáneamente tanto a L_1 como a L_2.

Observemos que la suma de los subespacios, como también la intersección de ellos, siempre son conjuntos no vacíos, ya que les pertenece a ciencia cierta el vector nulo del espacio K. Demostremos que estos conjuntos son subespacios.

En efecto, tomemos dos vectores arbitrarios z_1, z_2 de la suma $L_1 + L_2$. Esto significa que $z_1 = x_1 + y_1$, $z_2 = x_2 + y_2$, donde $x_1, x_2 \in L_1$ e $y_1, y_2 \in L_2$. Examinaremos ahora una combinación lineal arbitraria $\alpha z_1 + \beta z_2$. Tenemos $\alpha z_1 + \beta z_2 = (\alpha x_1 + \beta x_2) + (\alpha y_1 + \beta y_2)$. Puesto que $\alpha x_1 + \beta x_2 \in L_1$ y $\alpha y_1 + \beta y_2 \in L_2$, entonces $\alpha z_1 + \beta z_2 \in L_1 + L_2$. Por consiguiente, $L_1 + L_2$ es un subespacio. Sea, ahora, $z_1, z_2 \in L_1 \cap L_2$, es decir, $z_1, z_2 \in L_1$ y $z_1, z_2 \in L_2$. Está claro que $\alpha z_1 + \beta z_2 \in L_1$ y $\alpha z_1 + \beta z_2 \in L_2$, es decir, $\alpha z_1 + \beta z_2 \in L_1 \cap L_2$. Por consiguiente, $L_1 \cap L_2$ es también un subespacio.

De este modo, las operaciones de adición de los subespacios y de su intersección son algebraicas. Estas operaciones son, evidentemente, conmutativas y asociativas. Además, para todo subespacio L de K se verifica

$$L + 0 = L, \quad L \cap K = L.$$

Las leyes distributivas, que ligan ambas operaciones, están ausentes.

Como se nota con facilidad ya en los ejemplos más sencillos, la dimensión de la suma de dos subespacios arbitrarios depende no

sólo de la dimensión de los propios subespacios, sino también de cuán grande es su parte común. Resulta válido el

TEOREMA 19.1. *Para cualesquiera dos subespacios L_1, L_2 de dimensión finita tiene lugar la igualdad*

$$\dim (L_1 \cap L_2) + \dim (L_1 + L_2) = \dim L_1 + \dim L_2. \qquad (19.1)$$

DEMOSTRACIÓN. Designaremos las dimensiones de los subespacios L_1, L_2, $L_1 \cap L_2$ con r_1, r_2, m, respectivamente. Elijamos en la intersección $L_1 \cap L_2$ una base cualquiera $c_1, \ldots, c_m$. Los vectores son linealmente independientes y se disponen en L_1. Conforme al lema 19.1, en L_1 existen tales vectores $a_1, \ldots, a_k$ que el sistema $a_1, \ldots, a_k, c_1, \ldots, c_m$ será la base en L_1. Por analogía, en el subespacio L_2 existen tales vectores $b_1, b_2, \ldots, b_p$, que el sistema $b_1, \ldots, b_p, c_1, \ldots, c_m$ será la base en L_2. En este caso

$$r_1 = k + m, \qquad r_2 = p + m.$$

Si demostramos que el sistema de vectores

$$a_1, \ldots, a_k, \quad c_1, \ldots, c_m, \quad b_1, \ldots, b_p \qquad (19.2)$$

es la base del subespacio $L_1 + L_2$, entonces la afirmación del teorema tiene lugar, puesto que

$$m + (k + m + p) = (k + m) + (p + m).$$

Todo vector de los subespacios L_1, L_2 se expresa linealmente en términos de los vectores de su base y con mayor razón, linealmente en términos de los vectores (19.2). Por eso en términos de estos vectores también se expresará linealmente cualquier vector de la suma $L_1 + L_2$. Resta mostrar que el sistema (19.2) es linealmente independiente. Sea

$$\alpha_1 a_1 + \ldots + \alpha_k a_k + \gamma_1 c_1 + \ldots + \gamma_m c_m + \beta_1 b_1 + \ldots + \beta_p b_p = 0. \qquad (19.3)$$

Denotemos

$$b = \beta_1 b_1 + \ldots + \beta_p b_p. \qquad (19.4)$$

Está claro que $b \in L_1$. Pero de (19.3) se deduce que $b \in L_2$. Por consiguiente, $b \in L_1 \cap L_2$, es decir,

$$b = \nu_1 c_1 + \ldots + \nu_m c_m \qquad (19.5)$$

para ciertos números $\nu_1, \ldots, \nu_m$. Al comparar (19.4), (19.5) obtenemos

$$\beta_1 b_1 + \ldots + \beta_p b_p + (-\nu_1) c_1 + \ldots + (-\nu_m) c_m = 0.$$

El sistema de vectores $b_1, \ldots, b_p, c_1, \ldots, c_m$ es linealmente independiente por construcción y por ello

$$\beta_1 = \ldots = \beta_p = \nu_1 = \ldots = \nu_m = 0.$$

En virtud de la independencia lineal del sistema de vectores $a_1, \ldots, a_k, b_1, \ldots, c_m$ de (19.3) se desprende ahora que

$$\alpha_1 = \ldots = \alpha_k = \gamma_1 = \ldots = \gamma_m = 0.$$

El teorema queda demostrado.

Ejercicios.

1. Establézcase, en el ejemplo del espacio lineal V_3, el significado geométrico de las operaciones de la suma e intersección de subespacios.
2. ¿Qué es la suma de los subespacios V_1 y V_2?
3. ¿Qué es la intersección de los subespacios V_1 y V_2?
4. Demuéstrese que la dimensión de una intersección de cualquier número de subespacios no es superior a la mínima de las dimensiones de estos subespacios.
5. Demuéstrese que la dimensión de una suma de cualquier número de subespacios no es inferior a la máxima de las dimensiones de estos subespacios.

§ 20. Suma directa de los subespacios

Sean dados los subespacios $L_1, L_2, \ldots, L_m$ de cierto espacio lineal. Por definición de la operación de adición, todo vector z, perteneciente a la suma

$$K = L_1 + L_2 + \ldots + L_m, \tag{20.1}$$

puede ser representado en la forma

$$z = z_1 + z_2 + \ldots + z_m, \tag{20.2}$$

donde $z_i \in L_i$ para i cualquiera. En el caso general esta representación no será única. En cambio, si todo vector de K admite una única representación (20.2), entonces la suma (20.1) se llama *suma directa* y se designa del modo siguiente:

$$K = L_1 \dotplus L_2 \dotplus \ldots \dotplus L_m. \tag{20.3}$$

Las sumas directas poseen varias propiedades específicas. No obstante, para nosotros serán de interés en tanto dichas propiedades como *los rasgos comunes en la descomposición* (20.2) *y la descomposición según la base*. Supongamos que cierto espacio K puede descomponerse en suma directa (20.3) de sus subespacios $L_1, L_2, \ldots, L_m$. Entonces, debido a la unicidad de la descomposición (20.2), el sistema de los subespacios $L_1, L_2, \ldots, L_m$ puede considerarse como cierta "base generalizada" del espacio K y la descomposición (20.2), como descomposición según la "base generalizada". Tal interpretación de la suma directa es especialmente útil al estudiar los espacios lineales de gran dimensión, puesto que en estos espacios hemos de estudiar, como regla, no todas las componentes de la descomposición según la base, sino sólo una pequeña parte de ellas. El uso de la suma directa

hace posible evitar tanto las descomposiciones voluminosas como las investigaciones de los pormenores superfluos.

Sea K un espacio lineal n-dimensional. Tomemos su base arbitraria $e_1, e_2, \ldots, e_n$ y construyamos una totalidad de cápsulas lineales $L_1 = L_1(e_1)$, $L_2 = L_2(e_2)$, ..., $L_n = L_n(e_n)$. Es evidente en este caso que K será la suma directa de estos n subespacios unidimensionales. Pero el espacio K puede ser representado por diferentes medios en forma de una suma directa de los subespacios que tengan otra dimensión. El fundamento para tal representación lo constituye el

TEOREMA 20.1. *Para que el espacio K sea una suma directa de sus subespacios $L_1, \ldots, L_m$, es necesario y suficiente que la reunión de las bases de estos subespacios constituya la base de todo el espacio.*

DEMOSTRACIÓN. Supongamos que K es una suma directa de los subespacios $L_1, \ldots, L_m$ y los vectores $e_1, \ldots, e_{s_1}; \ldots$ $\ldots; e_{s_{m-1}+1}, \ldots, e_{s_m}$ constituyen las bases de dichos subespacios. Entonces para cada vector x de K tiene lugar la descomposición (20.2). Al representar cada uno de los vectores x_i en forma de una descomposición según la base del subespacio correspondiente L_i, obtenemos que

$$x = \alpha_1 e_1 + \ldots + \alpha_{s_1} e_{s_1} + \ldots + \alpha_{s_{m-1}+1} e_{s_{m-1}+1} + \ldots + \alpha_{s_m} e_{s_m} \qquad (20.4)$$

para ciertos números $\alpha_1, \ldots, \alpha_{s_m}$.

De este modo, todo vector de K se puede representar en forma de una combinación lineal de los vectores $e_1, \ldots, e_{s_m}$. Para que se pueda afirmar que estos vectores constituyen la base del espacio K, nos resta demostrar su independencia lineal. Examinaremos la igualdad

$$\beta_1 e_1 + \ldots + \beta_{s_1} e_{s_1} + \ldots + \beta_{s_{m-1}+1} e_{s_{m-1}+1} + \ldots + \beta_{s_m} e_{s_m} = 0 \qquad (20.5)$$

con los coeficientes numéricos $\beta_1, \ldots, \beta_{s_m}$ y designaremos

$$\begin{gathered} \beta_1 e_1 + \ldots + \beta_{s_1} e_{s_1} = y_1, \\ \ldots\ldots\ldots\ldots\ldots \\ \beta_{s_{m-1}+1} e_{s_{m-1}+1} + \ldots + \beta_{s_m} e_{s_m} = y_m. \end{gathered} \qquad (20.6)$$

Es obvio que $y_i \in L_i$ y de (20.5) proviene la siguiente igualdad:

$$0 = y_1 + \ldots + y_m.$$

Todos los subespacios contienen el vector nulo, por lo cual se verifica a ciencia cierta la correlación

$$0 = 0 + \ldots + 0.$$

Debido a la unicidad de la descomposición del vector nulo de K según los subespacios $L_1, \ldots, L_m$ concluimos que

$$y_1 = \ldots = y_m = 0.$$

De aquí se desprende que todos los coeficientes de las combinaciones lineales (20.6) son iguales a cero, es decir, los vectores $e_1, \ldots, e_{i_m}$ son linealmente independientes.

Supongamos ahora que los vectores $e_1, \ldots, e_{i_1}; \ldots; e_{i_{m-1}+1}, \ldots, e_{i_m}$, que constituyen las bases de los subespacios $L_1, \ldots, L_m$ forman la base de K. En este caso para todo vector x de K tiene lugar una única descomposición (20.4). Al denotar

$$\begin{gathered} \alpha_1 e_1 + \ldots + \alpha_{i_1} e_{i_1} = x_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \alpha_{i_{m-1}+1} e_{i_{m-1}+1} + \ldots + \alpha_{i_m} e_{i_m} = x_m \end{gathered} \tag{20.7}$$

resulta que para x existe al menos una descomposición (20.2). Todo vector x_i de (20.7) es una combinación lineal de los vectores básicos L_i. En virtud de la unicidad de la descomposición (20.4) para el vector x_i llegamos a la conclusión de que la descomposición (20.2) para él es también única. El teorema queda demostrado.

Ejercicios.

1. ¿En qué condiciones el espacio V_3 será la suma directa de sus subespacios V_1 y V_2?
2. ¿En qué condiciones el espacio V_3 será la suma directa de sus dos subespacios del tipo V_2?
3. ¿Podrá ser V_3 una suma directa de sus dos subespacios del tipo V_1? Si no, ¿por qué?
4. Demuéstrese que para que la suma (20.1) sea directa, es necesario y suficiente que la descomposición (20.2) sea única para el vector nulo.
5. Demuéstrese que para que la suma (20.1) sea directa, es necesario y suficiente que la intersección de cada uno de los subespacios L_i, $i = 1, \ldots, m$, con la suma de los demás contenga sólo el vector nulo.

§ 21. Isomorfismo de los espacios lineales

Consideraremos el conjunto de todos los espacios lineales definidos sobre un mismo campo P. Será natural preguntar en qué consiste la semejanza y la diferencia entre todos estos espacios.

Todo espacio lineal contiene en su descripción dos facetas esencialmente diferentes. En primer lugar, el espacio lineal es una totalidad de objetos concretos que se denominan vectores. En segundo lugar, sobre estos objetos concretos están definidas las operaciones de adición y multiplicación por un número que poseen ciertas propiedades. Por esto, el interés puede dirigirse o bien hacia la naturaleza de los vectores y las propiedades de éstos o bien hacia las propiedades de las operaciones indicadas independientemente de la naturaleza de los elementos.

La naturaleza de los vectores nos interesaba sólo en el estudio de los segmentos dirigidos y únicamente en la medida que fue necesario para introducir operaciones y establecer las propiedades de las mismas. A continuación, la investigación posterior de los segmentos dirigidos se apoyaba exclusivamente en las propiedades de las operaciones. Del modo análogo procederemos también en cada caso concreto. Por ello, se considerará que dos espacios de una misma estructura respecto a las operaciones de adición y multiplicación por un número poseen propiedades iguales o son *isomorfos*. Con más precisión:

Dos espacios lineales dados sobre un mismo campo se llaman isomorfos, si se puede establecer tal correspondencia biunívoca entre sus vectores que a la suma de cualesquiera dos vectores del primer espacio le corresponda la suma de los vectores correspondientes del segundo espacio y al producto de un número por un vector del primer espacio le corresponda el producto de este mismo número por el vector correspondiente del segundo espacio.

Sean K y K' los espacios isomorfos. El hecho de que a todo vector x de K se le ha puesto en correspondencia un vector determinado x' de K' puede entenderse como introducción de cierta "función"

$$x' = \omega (x), \tag{21.1}$$

cuyo "argumento" es el vector x del espacio K y el "valor" es el vector x' del espacio K'. Ambas propiedades de esta función se pueden escribir del modo siguiente. Para cualesquiera x, y de K y para todo número λ

$$\omega (x + y) = \omega (x) + \omega (y),$$
$$\omega (\lambda x) = \lambda\omega (x). \tag{21.2}$$

El carácter biunívoco de la correspondencia entre K y K' significa que a cualesquiera argumentos diferentes de la función (21.1) les corresponden diferentes valores, es decir, si

$$x \neq y, \tag{21.3}$$

entonces

$$\omega (x) \neq \omega (y). \tag{21.4}$$

Por consiguiente, de la igualdad o desigualdad de los valores de la función proviene, correspondientemente, la igualdad o desigualdad de los argumentos.

Los espacios isomorfos tienen mucho en común. En particular, al vector nulo le corresponde el vector nulo, pues

$$\omega (0) = \omega (0 \cdot x) = 0 \cdot \omega (x) = 0 \cdot x' = 0'.$$

Sin embargo, el corolario más importante consiste en que a un sistema linealmente independiente de vectores le corresponde de nuevo un sistema linealmente independiente.

Efectivamente, sean $x_1, x_2, \ldots, x_n$ unos vectores linealmente independientes. Consideraremos una combinación lineal $\alpha_1 \omega(x_1) + \alpha_2 \omega(x_2) + \ldots + \alpha_n \omega(x_n)$ y hagámosla igual a cero. Debido a las propiedades de la correspondencia isomorfa tenemos

$$0' = \alpha_1 \omega(x_1) + \alpha_2 \omega(x_2) + \ldots + \alpha_n \omega(x_n) = \omega(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n) = \omega(0),$$

de donde se desprende que

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = 0.$$

Como los vectores $x_1, x_2, \ldots, x_n$ son linealmente independientes, todos los coeficientes deben ser nulos.

El corolario demostrado permite afirmar que si dos espacios lineales de dimensión finita son isomorfos, tienen la *misma* dimensión. La afirmación recíproca es también válida. A saber, tiene lugar el

TEOREMA 2.1. *Dos espacios lineales cualesquiera que tienen una misma dimensión finita y están dados sobre un mismo campo son isomorfos.*

DEMOSTRACIÓN. Sean K y K' dos espacios lineales de dimensión n. Elijamos una base $e_1, e_2, \ldots, e_n$ en el espacio K y una base $e_1', e_2', \ldots, e_n'$ en el espacio K'. Construyamos un isomorfismo ω del modo siguiente, haciendo uso de los sistemas indicados de vectores. A todo vector

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n$$

del espacio K le pondremos en correspondencia el vector

$$\omega(x) = \alpha_1 e_1' + \alpha_2 e_2' + \ldots + \alpha_n e_n'$$

del espacio K'. La correspondencia establecida será biunívoca, puesto que la descomposición según la base es única.

Elijamos ahora dos vectores arbitrarios x e y de K y un número cualquiera λ y supongamos que

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$
$$y = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n.$$

Se tiene

$$\omega(x + y) = \omega((\alpha_1 + \beta_1) e_1 + (\alpha_2 + \beta_2) e_2 + \ldots + (\alpha_n + \beta_n) e_n) =$$
$$= (\alpha_1 + \beta_1) e_1' + (\alpha_2 + \beta_2) e_2' + \ldots + (\alpha_n + \beta_n) e_n' =$$
$$= (\alpha_1 e_1' + \alpha_2 e_2' + \ldots + \alpha_n e_n') + (\beta_1 e_1' + \beta_2 e_2' + \ldots + \beta_n e_n') = \omega(x) + \omega(y).$$
$$\omega(\lambda x) = \omega((\lambda\alpha_1) e_1 + (\lambda\alpha_2) e_2 + \ldots + (\lambda\alpha_n) e_n) =$$
$$= (\lambda\alpha_1) e_1' + (\lambda\alpha_2) e_2' + \ldots + (\lambda\alpha_n) e_n' =$$
$$= \lambda(\alpha_1 e_1' + \alpha_2 e_2' + \ldots + \alpha_n e_n') = \lambda\omega(x).$$

Las igualdades obtenidas demuestran la validez de la afirmación del teorema.

La importancia de este teorema es muy grande. Precisamente este teorema permite ahora hablar con toda la certeza de que desde el punto de vista de todos los corolarios procedentes de los axiomas cualesquiera dos espacios lineales que tienen una misma dimensión y están dados sobre un mismo campo son indistinguibles. Por consiguiente, podríamos construir un espacio lineal n-dimensional sobre el campo dado y, al investigar sólo dicho espacio, aclarar las leyes propias a todos los espacios de dimensión finita.

Sea dado un campo P. Consideremos un conjunto cuyos elementos son toda clase de sistemas ordenados de n números $\alpha_1, \alpha_2, \ldots, \alpha_n$ del campo P. Si x es un elemento de este conjunto, escribiremos

$$x = (\alpha_1, \alpha_2, \ldots, \alpha_n). \tag{21.5}$$

Las operaciones de adición y multiplicación por el número λ del campo P determinemos del modo siguiente:

$$(\alpha_1, \alpha_2, \ldots, \alpha_n) + (\beta_1, \beta_2, \ldots, \beta_n) = \\ = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n), \tag{21.6}$$
$$\lambda(\alpha_1, \alpha_2, \ldots, \alpha_n) = (\lambda\alpha_1, \lambda\alpha_2, \ldots, \lambda\alpha_n).$$

Es fácil comprobar que los axiomas del espacio lineal están cumplidos. En particular, el vector nulo se determina por el sistema de ceros, es decir,

$$0 = (0, 0, \ldots, 0),$$

y el vector opuesto para el vector (21.5) será

$$-x = (-\alpha_1, -\alpha_2, \ldots, -\alpha_n).$$

Este espacio es n-dimensional y una de sus bases se indica con facilidad inmediatamente. A saber,

$$\begin{aligned} e_1 &= (1, 0, \ldots, 0, 0), \\ e_2 &= (0, 1, \ldots, 0, 0), \\ &\ldots\ldots\ldots\ldots \\ e_n &= (0, 0, \ldots, 0, 1). \end{aligned} \tag{21.7}$$

Puesto que para el elemento (21.5) tiene lugar la descomposición

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$

los números $\alpha_1, \alpha_2, \ldots, \alpha_n$ se llamarán *coordenadas* del vector x.

El espacio de tal tipo lo denominaremos *aritmético* y lo designaremos mediante P_n subrayando con esto su relación con el campo P. En el caso de que P sea un campo de los números complejos, reales o racionales, tales espacios n-dimensionales se designarán con C_n, R_n o D_n, respectivamente.

Ahora puede parecer que no hay ninguna necesidad en el estudio de los espacios lineales n-dimensionales arbitrarios. En efecto, sabemos que desde el punto de vista de los corolarios procedentes de los axiomas, los espacios lineales isomorfos son indistinguibles, razón por la cual siempre podemos estudiar con todo éxito solamente, por ejemplo, P_n. No obstante, los razonamientos generales permiten poner en evidencia las propiedades más importantes de los espacios lineales, es decir, aquellas que *no dependen de los sistemas básicos* o, en otras palabras, son *invariantes en los isomorfismos*.

Al estudiar solamente los espacios P_n, siempre estamos atados a *una base concreta*, por lo cual no podríamos ver con facilidad la invariación de unas u otras deducciones. Con todo eso, se debe observar que las propiedades particulares del espacio P_n no se mezclen con las propiedades generales de los espacios lineales. No es tan fácil de hacerlo ni mucho menos.

Para concluir, indiquemos una circunstancia más. Por analogía con el espacio P_n consideraremos el espacio P_∞ cuyos elementos son toda clase de surtidos infinitos ordenados de los números $\alpha_1, \alpha_2, \ldots$ del campo P. El elemento x de este conjunto lo designaremos por analogía con (21.5)

$$x = (\alpha_1, \alpha_2, \ldots)$$

y por analogía con (21.6) introduciremos las operaciones sobre los elementos.

El espacio P_∞ ya será de dimensión infinita. Si suponemos que los espacios de dimensión infinita son isomorfos respecto del espacio P_∞, no será difícil comprender que los espacios de dimensión infinita y los de dimensión finita han de tener mucho en común. De eso no se debe olvidar.

Ejercicios.

1. Constrúyase una correspondencia isomorfa entre el espacio V_1 y un espacio de números reales dado sobre un campo de números reales.
2. Constrúyase una correspondencia isomorfa entre el espacio V_2 y un espacio de números complejos dado sobre un campo de números reales.
3. Demuéstrese que en los espacios isomorfos los sistemas equivalentes de vectores se transforman en sistemas equivalentes.
4. Demuéstrese que en los espacios isomorfos una intersección de subespacios se transforma en otra intersección de subespacios.
5. Demuéstrese que en los espacios isomorfos una suma directa de subespacios se transforma en otra suma directa de subespacios

§ 22. Dependencia lineal y sistemas de ecuaciones lineales

La investigación de varias cuestiones, ligadas de uno u otro modo con la dependencia lineal, se reduce a la resolución del problema siguiente.

Sean dados un sistema de vectores $a_1, a_2, \dots, a_m$ y el vector b. Se necesita establecer si el vector b es una combinación lineal del sistema dado de vectores y, además, hallar los coeficientes de la combinación lineal.

Si el vector b es una combinación lineal de los vectores $a_1, a_2, \dots, a_m$, entonces existen tales números $x_1, x_2, \dots, x_m$ que

$$x_1a_1 + x_2a_2 + \dots + x_ma_m = b. \qquad (22.1)$$

Por consiguiente, el problema planteado se reduce a la investigación de la ecuación vectorial (22.1) respecto de los números $x_1, x_2, \dots, x_m$.

Supongamos que los vectores se han dado por sus coordenadas respecto de cierta base $e_1, e_2, \dots, e_k$, es decir,

$$\begin{aligned} a_1 &= (a_{11}, a_{21}, \dots, a_{k1}), \\ a_2 &= (a_{12}, a_{22}, \dots, a_{k2}), \\ &\dots\dots\dots\dots \\ a_m &= (a_{1m}, a_{2m}, \dots, a_{km}), \\ b &= (b_1, b_2, \dots, b_k). \end{aligned}$$

Al igualar las coordenadas correspondientes de los vectores en los miembros primero y segundo de la ecuación (22.1), obtenemos

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m &= b_2, \\ &\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 + \dots + a_{km}x_m &= b_k. \end{aligned} \qquad (22.2)$$

Este sistema de ecuaciones que refleja la inscripción de la ecuación (22.1) en coordenadas se llama *sistema de las ecuaciones algebraicas lineales*. Los números $b_1, b_2, \dots, b_k$ se denominan *miembros derechos* y $x_1, x_2, \dots, x_m$, *incógnitas* del sistema de ecuaciones. La totalidad ordenada de los valores de las incógnitas que satisface cada una de las ecuaciones (22.2) recibe el nombre de *solución* del sistema. Si el sistema de ecuaciones algebraicas lineales tiene al menos una sola solución, se llama *compatible*; en el caso contrario el sistema se denomina *incompatible*.

De este modo, la respuesta a la pregunta sobre si es o no el vector b una combinación lineal de los vectores $a_1, a_2, \dots, a_m$ se determina por la compatibilidad o no compatibilidad del sistema (22.2). Si el sistema es compatible, cualquier solución de él proporciona los coeficientes de la descomposición del vector b según el sistema de vectores $a_1, a_2, \dots, a_m$.

Dos sistemas de ecuaciones algebraicas lineales respecto de las mismas incógnitas se llaman *equivalentes*, si toda solución de un

sistema es solución para el otro sistema o ambos sistemas son incompatibles.

El método general para solucionar los sistemas puede fundamentarse sobre la transformación sucesiva del sistema inicial (22.2) en tal sistema equivalente, para el cual la solución se halle de manera suficientemente simple. Demos a conocer ahora uno de los métodos llamado *método de eliminación* o *método de Gauss*.

El proceso de resolución incluye, en el caso general, no más que $k-1$ etapas. Con el fin de distinguir los coeficientes de las incógnitas y los miembros derechos, que se obtienen en el transcurso de la transformación en diferentes etapas, marcarémoslos con un índice adicional que se pone por arriba y significa el número de la etapa realizada. De conformidad con esta observación, el sistema inicial (22.2) tendrá la forma

$$\begin{aligned} a_{11}^{(0)} z_1 + a_{12}^{(0)} z_2 + \ldots + a_{1m}^{(0)} z_m &= b_1^{(0)}, \\ a_{21}^{(0)} z_1 + a_{22}^{(0)} z_2 + \ldots + a_{2m}^{(0)} z_m &= b_2^{(0)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{k1}^{(0)} z_1 + a_{k2}^{(0)} z_2 + \ldots + a_{km}^{(0)} z_m &= b_k^{(0)}. \end{aligned} \qquad (22.3)$$

Examinemos la primera ecuación. Si todos los coeficientes de las incógnitas y el miembro derecho son iguales a cero, entonces cualquier totalidad de los números $z_1, z_2, \ldots, z_m$ satisfará esta ecuación. Por consiguiente, al excluir de la consideración la primera ecuación, obtendremos un sistema equivalente. Puede suceder que todos los coeficientes de las incógnitas en la primera ecuación son iguales a cero, mientras que el miembro derecho no es nulo. En este caso ninguna totalidad de los números $z_1, z_2, \ldots, z_m$ puede satisfacer tal ecuación. Entonces, el sistema será incompatible y la investigación se ha dado por terminada.

Supongamos que entre los coeficientes de las incógnitas en la primera ecuación se tiene aunque uno diferente de cero. Sin restringir la generalidad, se puede considerar que $a_{11}^{(0)} \neq 0$, puesto que en el caso contrario se lo puede conseguir por permutación de las incógnitas. Llamemos el elemento $a_{11}^{(0)}$ elemento *rector*. Expresemos la incógnita z_1 de la primera ecuación en términos de las incógnitas restantes y del miembro derecho y a continuación sustituyamos la expresión obtenida para z_1 en todas las ecuaciones, salvo la primera. Reduciendo en todas las ecuaciones los términos semejantes obtendremos un sistema nuevo

$$\begin{aligned} a_{11}^{(0)} z_1 + a_{12}^{(0)} z_2 + \ldots + a_{1m}^{(0)} z_m &= b_1^{(0)}, \\ a_{22}^{(1)} z_2 + \ldots + a_{2m}^{(1)} z_m &= b_2^{(1)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{k2}^{(1)} z_2 + \ldots + a_{km}^{(1)} z_m &= b_k^{(1)}. \end{aligned} \qquad (22.4)$$

Los coeficientes de este sistema están ligados con los del sistema anterior mediante las siguientes correlaciones:

$$a_{ij}^{(1)} = a_{ij}^{(0)} - a_{1j}^{(0)} \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}$$

$$b_i^{(1)} = b_i^{(0)} - b_1^{(0)} \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}$$

para cualesquiera i, j.

Los sistemas (22.3) y (22.4) son equivalentes. En efecto, supongamos que el sistema (22.3) es compatible. Entonces, cualquier solución $x_1, x_2, \ldots, x_m$ del mismo convierte todas las ecuaciones del sistema (22.3) en identidades. Reiterando de nuevo con cualquiera de las soluciones el proceso de eliminación, nos convencemos de que la solución mencionada es también solución del sistema (22.4). Supongamos luego que cierta solución del sistema (22.4) no es la solución del sistema (22.3). Satisface a ella cierta la primera ecuación de (22.3). Supongamos que dicha solución no satisface una ecuación de número $i \geqslant 2$. Entonces, repitiendo otra vez el proceso de eliminación, llegamos a que la solución elegida no debe satisfacer la i-ésima ecuación del sistema (22.4). Pero esto contradice a la hipótesis. Ahora está claro que si uno de los sistemas es incompatible, el otro sistema será también incompatible.

Hemos descrito solamente la primera etapa de la transformación del sistema. Todas las etapas restantes se realizan de la manera análoga. En la segunda etapa eliminamos la incógnita x_2 de todas las ecuaciones a excepción de las dos primeras; en la tercera etapa, la incógnita x_3 de todas las ecuaciones a excepción de las tres primeras, etc. Si, en el transcurso de las transformaciones, no encontramos las ecuaciones, donde todos los coeficientes de las incógnitas son iguales a cero, entonces acabadas $k-1$ etapas, llegaremos al sistema:

$$\begin{array}{l|l}
a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + \ldots + a_{1k}^{(0)}x_k & + a_{1,k+1}^{(0)}x_{k+1} + \ldots + a_{1m}^{(0)}x_m = b_1^{(0)} \\
\qquad a_{22}^{(1)}x_2 + \ldots + a_{2k}^{(1)}x_k & + a_{2,k+1}^{(1)}x_{k+1} + \ldots + a_{2m}^{(1)}x_m = b_2^{(1)} \\
\ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots\ldots \\
\qquad\qquad a_{kk}^{(k-1)}x_k & + a_{k,k+1}^{(k-1)}x_{k+1} + \ldots + a_{km}^{(k-1)}x_m = b_k^{(k-1)},
\end{array} \qquad (22.5)$$

que es equivalente al sistema (22.4). Si en el transcurso de la transformación encontramos unas ecuaciones que se satisfacen idénticamente, el sistema (22.5) se compondrá de un número menor de ecuaciones.

Las incógnitas $x_{k+1}, \ldots, x_n$ se denominan *indeterminadas*. Evidentemente, cualesquiera que sean los valores que se atribuyan a estas incógnitas, todas las demás incógnitas del sistema (22.5), a partir de x_k, también pueden determinarse sucesivamente. Los coeficientes $a_{11}^{(0)}, a_{22}^{(1)}, \ldots, a_{kk}^{(k-1)}$, por los cuales tendremos que dividir, son elementos rectores en las etapas separadas, razón por la cual todos ellos son distintos de cero.

Así pues, desde el punto de vista *teórico*, el concepto de dependencia lineal se ha investigado en un grado suficientemente completo. No obstante, en el aspecto *práctico* puede conducir a unas dificultades muy serias. Por ejemplo, consideraremos en el espacio R_k un sistema de vectores

$$\begin{aligned} a_1 &= (1, \quad -2, \quad 0, \ldots, 0, \quad 0),\\ a_2 &= (0, \quad 1, \quad -2, \ldots, 0, \quad 0),\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ a_{k-1} &= (0, \quad 0, \quad 0, \ldots, 1, \quad -2),\\ a_k &= (-2^{-(k-1)}, \quad 0, \quad 0, \ldots, 0, \quad 1). \end{aligned} \qquad (22.6)$$

Es linealmente dependiente, puesto que

$$2^{-k}a_1 + 2^{-(k-1)}a_2 + \ldots + 2^{-1}a_k = 0.$$

Observemos que $2^{-(k-1)} < 10^{-12}$ para $k > 40$, por lo cual al realizar los cálculos prácticos, surge, naturalmente, el deseo de menospreciar el valor de la coordenada tan pequeña. Además, todos los números, como regla, suelen conocerse con inexactitud y casi siempre contienen errores mucho más grandes. Pero, si incluso conociéramos las coordenadas con toda la exactitud, ya los cálculos iniciales con ellas llevarían a los resultados inexactos, si los cálculos se realizaban de manera aproximada. Se debe añadir que la mayoría de las máquinas computadoras modernas no percibe los números tan pequeños como es $2^{-(k-1)}$ para $k > 64$, y opera con ellos como si fueran ceros. Por esta razón, en realidad, en lugar del sistema de vectores (22.6) en la práctica podemos tropezar con un sistema del tipo:

$$\begin{aligned} a_1 &= (1, \quad -2, \quad 0, \ldots, 0, \quad 0),\\ a_2 &= (0, \quad 1, \quad -2, \ldots, 0, \quad 0),\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ a_{k-1} &= (0, \quad 0, \quad 0, \ldots, 1, \quad -2),\\ a_k &= (0, \quad 0, \quad 0, \ldots, 0, \quad 1). \end{aligned} \qquad (22.7)$$

Mas, este sistema es linealmente independiente.

De este modo, las alteraciones pequeñas en coordenadas de los vectores pueden llevar a que bajo las condiciones en que las propias

coordenadas se prefijen de modo aproximado y los cálculos con ellas sean también aproximados, un sistema linealmente dependiente puede hacerse linealmente independiente y, viceversa, un sistema linealmente independiente, linealmente dependiente. Pero en este caso será natural preguntar ¿qué sentido práctico tienen los conceptos de dependencia lineal, rango, base, sistemas compatible y no compatible y, en general, todo lo que hemos investigado hasta ahora? Para esta pregunta no existe respuesta sencilla, puesto que está asociada con una comprensión profunda de los problemas que se resuelven. *Esta pregunta da origen a las diferencias que distinguen la matemática "precisa" de la matemática "aproximada".*

Ejercicios.

1. Demuéstrese que si el sistema (22.2) es compatible, tendrá una única solución cuando, y sólo cuando, el sistema de vectores $a_1, a_2, \ldots, a_n$ sea linealmente independiente.

2. Demuéstrese que si el sistema de vectores $a_1, a_2, \ldots, a_m$ tiene el rango r, el sistema (22.5) se compondrá de r ecuaciones.

3. Como soluciones del sistema consideraremos los vectores del espacio P_m. Supongamos que $b = 0$ y el sistema de vectores $a_1, a_2, \ldots, a_m$ tiene el rango r. Demuéstrese que el conjunto de todas las soluciones del sistema (22.2) forma en este caso un subespacio $(n - r)$-dimensional del espacio P_m.

4. Hállense todas las soluciones del sistema de ecuaciones algebraicas lineales

$$\sqrt{2}\,x_1 + 1 \cdot x_2 = \sqrt{3},$$
$$2\,x_1 + \sqrt{2}\,x_2 = \sqrt{6}.$$

Resuélvase este mismo sistema al prefijar $\sqrt{2}$, $\sqrt{3}$, $\sqrt{6}$ con la exactitud diferente. Compárense los resultados entre sí.

5. Establézcase la relación existente entre el método de Gauss y las transformaciones elementales de un sistema de vectores.

CAPÍTULO 3 MEDICIONES EN EL ESPACIO LINEAL

§ 23. Sistemas afines de coordenadas

Hay gran cantidad de los problemas científicos y técnicos que requieren descripción exacta de la posición en el espacio de diferentes objetos geométricos tales como el punto, la figura, la línea, la superficie, etc. Cuando se trata de un objeto complejo, resulta muy importante conocer no sólo la característica general de la posición, como, por ejemplo, la ubicación del centro de gravedad, sino también la posición de todo punto del objeto.

A título de ejemplo recordemos que la predicción de los eclipses lunares y solares es posible gracias a que se conoce la posición de los cuerpos celestes en cualquier momento de tiempo. La transmisión de las imágenes de televisión a grandes distancias puede realizarse porque la posición de todo punto de la imagen que se transmite está bien definida.

Evidentemente, es necesario proporcionar un método que describa la posición de un solo punto, pues todo objeto geométrico puede ser definido como cierta totalidad de puntos. En este caso, obviamente conviene considerar independientemente la posición de un punto en una recta, en un plano o en un espacio, puesto que la descripción espacial de un objeto es lejos de ser siempre oportuno. Por ejemplo, una losa puede considerarse, a ciencia cierta, sólo en un plano, mientras el movimiento de un punto material, siendo ausentes las fuerzas que actúan contra él, sólo en una línea recta.

Una de las descripciones más difundidas de la posición de un punto está basada sobre una idea muy simple. Ya hemos notado que entre todos los puntos y los segmentos dirigidos sujetados puede establecerse una correspondencia biunívoca. Por ello, la descripción de la posición de un punto puede ser sustituida por la descripción de la posición del segmento dirigido correspondiente. Esta última se determina plenamente por las coordenadas del segmento respecto de cualquier base, es decir, por ciertos surtidos ordenados de números. Por consiguiente, la posición de un punto debe determinarse también por los surtidos ordenados de números. Pasamos, ahora, a la investigación de esta idea.

Sea dada una línea recta. Fijemos en ésta un punto arbitrario O y consideremos un espacio lineal V_1 de vectores dispuestos en la recta dada y sujetados al punto O. Elijamos en el citado espacio un vector básico $\mathbf{a}$. Transformemos luego la recta en un eje, al definir en la misma una dirección de modo tal que la magnitud del segmento $\mathbf{a}$ sea positiva (fig. 23.1).

Un eje con el punto O y el vector básico $\mathbf{a}$, definidos en el eje, forma *un sistema afín de coordenadas en la línea recta*. El punto O



Fig. 23.1.

se llama *origen del sistema de coordenadas* y la longitud del vector $\mathbf{a}$, *unidad de escala*.

La posición de cualquier punto M en la recta se determina unívocamente por la posición del vector $\overrightarrow{OM}$. Los vectores $\mathbf{a}$, $\overrightarrow{OM}$ son colineales y $\mathbf{a} \neq 0$, a consecuencia de lo cual, de acuerdo con el corolario del lema 18.1, existe tal número real α que

$$\overrightarrow{OM} = \alpha \mathbf{a}. \tag{23.1}$$

Este número se denomina *coordenada afín* del punto M en la recta. El hecho de que el punto M tiene la coordenada α se denota por el símbolo $M(\alpha)$.

Hemos de notar que, siendo fijado un sistema afín de coordenadas en la recta, la correlación (23.1) define unívocamente la coordenada afín α de cualquier punto M en la recta. Evidentemente, lo recíproco es también cierto. A saber, todo número α determina unívocamente, mediante la correlación (23.1), un cierto punto M de la línea recta. De este modo, siendo fijado un sistema afín de coordenadas, *existe una correspondencia biunívoca entre todos los números reales y los puntos de la recta*.

La definición de los puntos por medio de sus coordenadas permite calcular las magnitudes de los segmentos dirigidos y las distancias entre los puntos. Sean dados los puntos $M_1(\alpha_1)$ y $M_2(\alpha_2)$. Tenemos

$$\{\overrightarrow{M_1M_2}\} = \{\overrightarrow{OM_2} - \overrightarrow{OM_1}\} = \{\alpha_2\mathbf{a} - \alpha_1\mathbf{a}\} =$$
$$= \{(\alpha_2 - \alpha_1)\mathbf{a}\} = (\alpha_2 - \alpha_1)\{\mathbf{a}\} = (\alpha_2 - \alpha_1)|\mathbf{a}|. \tag{23.2}$$

Al designar mediante $\rho(M_1, M_2)$ la distancia entre los puntos M_1 y M_2, tenemos

$$\rho(M_1, M_2) = |\{\overrightarrow{M_1M_2}\}| = |\alpha_2 - \alpha_1|\,|\mathbf{a}|. \tag{23.3}$$

Las fórmulas se hacen particularmente sencillas, si la longitud del vector básico es igual a la unidad. En este caso

$$\begin{aligned}(\overrightarrow{M_1M_2}) &= \alpha_2-\alpha_1,\\ \rho(M_1,\ M_2) &= |\alpha_2-\alpha_1|.\end{aligned} \tag{23.4}$$

Supongamos ahora que se ha dado un cierto plano. Fijemos en el mismo un punto arbitrario O y consideremos el espacio lineal V_2 de los vectores dispuestos en el plano dado y sujetados al punto O. Elijamos en el espacio un par cualquiera de vectores básicos a, b. A las líneas rectas que contienen dichos vectores atribuyámosles direcciones determinando de un modo tal que las magnitudes de los segmentos a, b sean positivas (fig. 23.2).

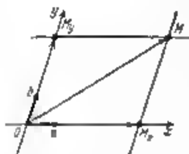


Fig. 23.2.

En un plano dos ejes que se cortan en el punto O y llevan los vectores básicos a, b dados forman *un sistema afín de coordenadas en el plano*. El eje que contiene el primer vector básico se llama eje Ox o *eje de abscisas*; el eje que contiene el segundo vector básico se denomina eje Oy o *eje de ordenadas*.
La posición de cualquier punto M en el plano se determina también unívocamente por el vector $\overrightarrow{OM}$ y para éste, a su vez, existe la única descomposición del tipo

$$\overrightarrow{OM}=\alpha a+\beta b. \tag{23.5}$$

Los números reales α, β se llaman de nuevo *coordenadas afines* del punto M. La primera coordenada se llama *abscisa* y la segunda, *ordenada* de M. El hecho de que el punto M tiene las coordenadas α, β se denota mediante el símbolo $M(\alpha,\ \beta)$.

En los ejes de coordenadas Ox, Oy existen los únicos puntos M_x, M_y tales que

$$\overrightarrow{OM}=\overrightarrow{OM_x}+\overrightarrow{OM_y}. \tag{23.6}$$

Dichos puntos se disponen en la intersección de los ejes de coordenadas con las rectas que son paralelas a los ejes y pasan por el punto M. Se denominan *proyecciones afines* del punto M sobre los ejes de coordenadas. Los vectores $\overrightarrow{OM_x}$, $\overrightarrow{OM_y}$ se llaman *proyecciones afines del vector* $\overrightarrow{OM}$. Debido a la unicidad de las descomposiciones (23.5),

(23.6) llegamos a la conclusión de que

$$\overrightarrow{OM_x}=\alpha a, \qquad \overrightarrow{OM_y}=\beta b. \tag{23.7}$$

De este modo, si el punto M tiene las coordenadas $M(\alpha, \beta)$ los puntos M_x, M_y, que pertenecen al plano, tienen las coordenadas $M_x(\alpha, 0)$, $M_y(0, \beta)$. Además, si

$$\overrightarrow{OM}=(\alpha, \beta),$$

entonces

$$\overrightarrow{OM_x}=(\alpha, 0), \qquad \overrightarrow{OM_y}=(0, \beta).$$

Todo vector básico forma en su eje un sistema propio de coordenadas. Por ello, los puntos M_x, M_y se pueden considerar también como puntos de los ejes Ox, Oy, dados en estos sistemas propios de coordenadas. Sin embargo, de (23.7) proviene que la coordenada del punto M_x en el eje Ox es igual a la abscisa del punto M. Análogamente, la coordenada del punto M_y en el eje Oy es igual a la ordenada del punto M. Siendo estas afirmaciones del todo evidentes, son de gran importancia, puesto que permiten emplear las fórmulas (23.2)—(23.4).

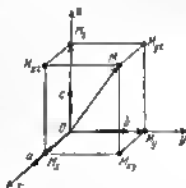


Fig. 23.3.

El par ordenado de números α, β determina unívocamente un punto. En efecto, las correlaciones (23.7) permiten construir unívocamente las proyecciones afines de un punto, las cuales, a su vez, determinan unívocamente un punto del plano. Por consiguiente, siendo fijado un sistema afín de coordenadas, *existe una correspondencia biunívoca entre todos los pares ordenados de números reales y puntos del plano*.

De modo análogo se introduce el sistema afín de coordenadas en un espacio. Fijemos un punto O y consideremos el espacio lineal V_3 de vectores sujetados al punto O. Elijamos en este espacio una terna de los vectores básicos a, b, c. Atribuyamos a las rectas que contienen dichos vectores las direcciones determinadas de modo tal que las magnitudes de los segmentos a, b, c sean positivas (fig. 23.3).

Tres ejes en el espacio que se cortan en el punto O y llevan definidos en sí los vectores básicos a, b, c, forman *un sistema afín de coordenadas en el espacio*. El eje que contiene el primer vector básico se llama eje Ox o eje de abscisas, el eje con el segundo vector básico es Oy o eje de ordenadas, el tercer eje se denomina Oz o eje de z-

coordenadas. Los ejes de coordenadas elegidos dos a dos determinan los así llamados *planos de coordenadas* que se denominarán Oxy, Oyz y Oxz.

La posición de cualquier punto M del espacio se determina, como antes, por el vector $\overrightarrow{OM}$, para el cual existe una única descomposición

$$\overrightarrow{OM} = \alpha a + \beta b + \gamma c.$$

Los números reales α, β, γ se llaman *coordenadas afines* del punto M en el espacio. La primera coordenada se llama *abscisa*, la segunda, *ordenada* y la coordenada tercera, *z-coordenada* del punto M. El hecho de que el punto M tiene las coordenadas α, β, γ se indicará mediante el símbolo $M(\alpha, \beta, \gamma)$.

Tracemos por el punto M del plano unos planos paralelos a los de coordenadas. Los puntos de intersección de estos planos con los ejes de coordenadas Ox, Oy, Oz se designarán con M_x, M_y, M_z, y se denominarán *proyecciones afines del punto M sobre los ejes de coordenadas*. La intersección de los planos de coordenadas con los pares de planos que pasan por el punto M, determina los puntos M_{yz}, M_{xz}, M_{xy}, que llevarán los nombres de *proyecciones afines del punto M sobre los planos de coordenadas*. Respectivamente, los vectores $\overrightarrow{OM}_{yz}$, $\overrightarrow{OM}_x$, etc. se denominarán *proyecciones afines del vector* $\overrightarrow{OM}$. Es evidente que

$$\overrightarrow{OM} = \overrightarrow{OM}_x + \overrightarrow{OM}_y + \overrightarrow{OM}_z,$$

$$\overrightarrow{OM}_{yz} = \overrightarrow{OM}_y + \overrightarrow{OM}_z,$$

$$\overrightarrow{OM}_{xz} = \overrightarrow{OM}_x + \overrightarrow{OM}_z,$$

$$\overrightarrow{OM}_{xy} = \overrightarrow{OM}_x + \overrightarrow{OM}_y.$$

Al igual que en el caso de un plano, si el punto M tiene las coordenadas

$$M(\alpha, \beta, \gamma),$$

las proyecciones afines de este punto tendrán por coordenadas:

$$M_x(\alpha, 0, 0), \quad M_y(0, \beta, 0), \quad M_z(0, 0, \gamma), \qquad (23.8)$$
$$M_{yz}(0, \beta, \gamma), \quad M_{xz}(\alpha, 0, \gamma), \quad M_{xy}(\alpha, \beta, 0).$$

Por analogía, si

$$\overrightarrow{OM} = (\alpha, \beta, \gamma),$$

se tiene

$$\overrightarrow{OM_x} = (\alpha,\ 0,\ 0), \quad \overrightarrow{OM_y} = (0,\ \beta,\ 0), \quad \overrightarrow{OM_z} = (0,\ 0,\ \gamma),$$

$$\overrightarrow{OM_{yz}} = (0,\ \beta,\ \gamma), \quad \overrightarrow{OM_{zx}} = (\alpha,\ 0,\ \gamma), \quad \overrightarrow{OM_{xy}} = (\alpha,\ \beta,\ 0).$$

Esta vez también, todo vector básico y todo par de vectores básicos forman sistemas afines propios en los ejes de coordenadas y en los planos de coordenadas. Las coordenadas de los puntos en estos sistemas coinciden nuevamente con las coordenadas afines de los mismos, considerados como puntos del espacio. Ahora bien, siendo fijado un sistema afín de coordenadas, *existe una correspondencia biunívoca entre todas las ternas ordenadas de números reales y los puntos del espacio.*

Entre los sistemas afines de coordenadas en una recta, en un plano y en un espacio, son de mayor uso los llamados *sistemas rectangulares cartesianos de coordenadas.* Estos últimos se caracterizan por el hecho de que todos los vectores básicos tienen una longitud igual a la unidad y los ejes de coordenadas, tanto en el caso de un plano como en el de un espacio, son perpendiculares dos a dos. Los vectores básicos en el sistema cartesiano de coordenadas se designan, comúnmente, por las letras i, j, k. En lo sucesivo haremos uso, como regla, sólo de los sistemas que acabamos de mencionar.

Ejercicios.

1. ¿Cuál de los puntos $A(\alpha)$, $B(-\alpha)$ se halla más a la derecha en el eje de coordenadas dibujado en la fig. 23.1?
2. ¿Qué representa en sí el lugar geométrico de los puntos $M(\alpha, \beta, \gamma)$ para los cuales las proyecciones M_{xy} tienen por coordenadas $M_{xy}(-3, 2, 0)$?
3. ¿Dependen las coordenadas de los puntos del modo de elegir la dirección en los ejes de coordenadas?
4. ¿Cómo varían las coordenadas de los puntos con el cambio de longitud de los vectores básicos?
5. ¿Qué coordenadas tiene el centro de un paralelepípedo, si el origen de coordenadas coincide con uno de sus vértices y los vectores básicos, con las aristas?

§ 24. Otros sistemas de coordenadas

Los sistemas de coordenadas empleados en las matemáticas permiten prefijar, mediante números, la posición de cualquier punto de un espacio, de un plano o de una recta. Esto hace posible toda clase de cálculos sobre las coordenadas y, lo que es muy importante, permite utilizar las computadoras modernas no sólo para diversos cálculos numéricos, sino también para la resolución de problemas geométricos, la investigación de cualesquiera objetos geométricos y correlaciones. Además de los

sistemas afines de coordenadas examinados anteriormente se emplean con frecuencia otros sistemas.

Sistema polar de coordenadas. Elijamos en un plano una recta y fijemos en la misma el sistema cartesiano de coordenadas. El origen O de este sistema se llamará *polo* y el eje de coordenadas, *eje polar*. Consideraremos en lo sucesivo que el segmento de escala del sistema de coordenadas en la recta se emplea para medir longitudes de los segmentos cualesquiera en el plano. Examinemos un punto arbitrario M del plano. Es evidente que su posición estará completamente definida, si se prefijan la distancia ρ entre los puntos M, O y el ángulo φ que se forma al hacer girar el rayo Ox alrededor del punto O en el *sentido contrahorario* hasta que su dirección coincida con la del segmento $\overrightarrow{OM}$ (fig. 24.1).

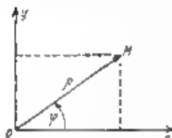


Fig. 24.1.

Se llaman *coordenadas polares* del punto M en un plano los números ρ y φ. El número ρ se llama *radio polar*, el número φ, *ángulo polar*. Comúnmente se supone que

$$0 \leqslant \rho < +\infty, \ 0 \leqslant \varphi < 2\pi. \tag{24.1}$$

Si el punto M coincide con el polo O, el ángulo polar se considera *indeterminado*.

Con cualquier sistema polar de coordenadas se enlaza de un modo natural cierto sistema rectangular cartesiano de coordenadas. En este sistema el origen de coordenadas coincide con el polo, el eje de abscisas, con el eje polar y el eje de ordenadas se obtiene girando el eje polar alrededor del punto O a un ángulo $\pi/2$.

Denotemos las coordenadas del punto M en el sistema rectangular cartesiano de coordenadas Oxy mediante α, β. Son evidentes las fórmulas

$$\alpha = \rho \cos \varphi, \quad \beta = \rho \operatorname{sen} \varphi.$$

De aquí obtenemos también las correlaciones inversas

$$\rho^2 = \alpha^2 + \beta^2, \ \cos \varphi = \frac{\alpha}{+(\alpha^2+\beta^2)^{1/2}},$$

$$\operatorname{sen} \varphi = \frac{\beta}{+(\alpha^2+\beta^2)^{1/2}}.$$

Estas fórmulas permiten calcular las coordenadas polares de un punto partiendo de sus coordenadas cartesianas, y viceversa.

Coordenadas cilíndricas. Elijamos en el espacio un plano cualquiera π y fijemos en éste un sistema polar de coordenadas. Por el polo O tracemos el eje Oz perpendicular al plano π (fig. 24.2). Coo-

vengamos, nuevamente, en considerar que para medir las longitudes de todos los segmentos en el espacio se utiliza un mismo segmento de escala. Introduzcamos en el plano π un sistema rectangular cartesiano de coordenadas que corresponde al sistema polar. Junto con el eje Oz el sistema introducido formará un sistema cartesiano de coordenadas en el espacio.

Consideremos las proyecciones M_z y M_{xy} del punto M sobre el eje Oz y el plano Oxy. El punto M_{xy}, como punto del plano π,

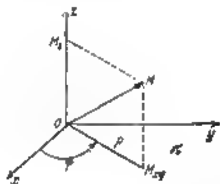


Fig. 24.2.

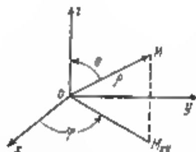


Fig. 24.3.

tiene por coordenadas polares ρ, φ. El punto M_z, que pertenece al eje Oz, tiene la coordenada z.

Se llaman *coordenadas cilíndricas* del punto M en el espacio tres números ρ, φ, z. En este caso se supone también que

$$0 \leqslant \rho < +\infty, \qquad 0 \leqslant \varphi < 2\pi.$$

El ángulo φ para los puntos del eje Oz *no está definido.*

La relación entre las coordenadas cartesianas en el sistema $Oxyz$ y las coordenadas cilíndricas se determina por medio de las correlaciones

$$x = \rho \cos \varphi, \qquad y = \rho \operatorname{sen} \varphi, \qquad z = z.$$

Coordenadas esféricas. Consideremos en un espacio el sistema rectangular cartesiano de coordenadas $Oxyz$ y el correspondiente sistema polar de coordenadas en el plano Oxy (fig. 24.3). Sea M un punto cualquiera del espacio distinto de O, y sea M_{xy} la proyección de M sobre el plano Oxy. Designemos con ρ la distancia del punto M al punto O y con θ, el ángulo formado por el vector $\overrightarrow{OM}$ y el vector básico del eje Oz. Sea, por fin, φ el ángulo polar de la proyección M_{xy}.

Se llaman *coordenadas esféricas* del punto M en el espacio tres números ρ, φ, θ. El número ρ es el *radio*, el número φ, *la longitud* y el número θ, *la latitud*. En este caso se supone que

$$0 \leqslant \rho < +\infty, \qquad 0 \leqslant \varphi < 2\pi, \qquad 0 \leqslant \theta \leqslant \pi.$$

La longitud no *está definida* para todos los puntos del eje Oz, mientras que la latitud no está definida para el punto O.

La relación existente entre las coordenadas cartesianas en el sistema $Oxyz$ y las coordenadas esféricas se determina por las correlaciones

$$x = \rho \operatorname{sen} \theta \cos \varphi, \qquad y = \rho \operatorname{sen} \theta \operatorname{sen} \varphi, \qquad z = \rho \cos \theta.$$

Ejercicios.

1. Constrúyase una línea tal que las coordenadas de sus puntos en el sistema polar satisfagan la correlación $\rho = \operatorname{sen} 3\varphi$.
2. Constrúyase una línea tal que las coordenadas de sus puntos en el sistema cilíndrico satisfagan las correlaciones $\rho = \varphi^{-1}$, $z = \varphi$.
3. Constrúyase una superficie tal que las coordenadas de sus puntos en el sistema esférico de coordenadas satisfagan las correlaciones

$$0 \leqslant \rho \leqslant 1, \qquad \varphi = \pi/2, \qquad 0 \leqslant \theta \leqslant \pi/2.$$

§ 25. Problemas

Consideremos varios problemas sencillos referentes a la aplicación de los sistemas rectangulares cartesianos de coordenadas. Para concretar, examinemos dichos problemas en el espacio. Los problemas análogos en un plano difieren de éstos sólo en pequeños detalles insignificantes. Convengamos en considerar siempre que se tiene un sistema fijado de coordenadas cuyo origen es el punto O y los vectores básicos son i, j, k.

Coordenadas de un vector. Supongamos que en el espacio están dados dos puntos $M_1(\alpha_1, \beta_1, \gamma_1)$ y $M_2(\alpha_2, \beta_2, \gamma_2)$. Estos puntos determinan el vector $\overrightarrow{M_1M_2}$, que tiene ciertas coordenadas respecto de la base i, j, k. Establezcamos la relación que existe entre las coordenadas del vector M_1M_2 y las de los puntos $\overleftrightarrow{M_1, M_2}$. Tenemos

$$\overrightarrow{M_1M_2} = \overrightarrow{OM_2} - \overrightarrow{OM_1}.$$

Luego, por definición de coordenadas afines de los puntos M_1, M_2,

$$\overrightarrow{OM_1} = \alpha_1 i + \beta_1 j + \gamma_1 k, \qquad \overrightarrow{OM_2} = \alpha_2 i + \beta_2 j + \gamma_2 k.$$

Por ello, de aquí se deduce que

$$M_1M_2 = (\alpha_2 - \alpha_1)\, i + (\beta_2 - \beta_1)\, j + (\gamma_2 - \gamma_1)\, k,$$

o bien, de acuerdo con las designaciones aceptadas,

$$\overrightarrow{M_1M_2} = (\alpha_2 - \alpha_1,\ \beta_2 - \beta_1,\ \gamma_2 - \gamma_1). \qquad (25.1)$$

Coordenadas de las proyecciones de un vector. Consideremos, otra vez, en el espacio el segmento dirigido $\overrightarrow{M_1M_2}$. Al proyectar los puntos M_1 y M_2 sobre el mismo plano de coordenadas o el mismo eje de coordenadas, obtendremos un nuevo segmento dirigido. Éste se llama *proyección coordenada del vector* $\overrightarrow{M_1M_2}$.

Todo vector de un espacio tiene seis proyecciones coordenadas: tres proyecciones sobre los ejes de coordenadas y tres sobre los planos de coordenadas. Son fáciles de calcular las coordenadas de las proyecciones en la base i, j, k según las coordenadas de los puntos M_1 $(\alpha_1,\ \beta_1,\ \gamma_1)$, M_2 $(\alpha_2,\ \beta_2,\ \gamma_2)$. Con este fin es suficiente hacer uso de las fórmulas (23.8), (25.1).

Supongamos, por ejemplo, que deseamos calcular las coordenadas de la proyección $\overrightarrow{M_{1x}M_{2x}}$. Teniendo presente que los puntos M_{1x} y M_{2x} tienen por coordenadas

$$M_{1x}\ (\alpha_1,\ 0,\ 0), \qquad M_{2x}\ (\alpha_2,\ 0,\ 0),$$

hallamos que

$$\overrightarrow{M_{1x}M_{2x}} = (\alpha_2 - \alpha_1,\ 0,\ 0). \qquad (25.2)$$

Por analogía,

$$\overrightarrow{M_{1xz}M_{2xz}} = (\alpha_2 - \alpha_1,\ 0,\ \gamma_2 - \gamma_1),$$

y así para el resto de las proyecciones.

Al comparar la primera de las fórmulas (23.4) con las fórmulas del tipo (25.2), llegamos a que

$$\{\overrightarrow{M_{1x}M_{2x}}\} = \alpha_2 - \alpha_1, \quad \{\overrightarrow{M_{1y}M_{2y}}\} = \beta_2 - \beta_1, \quad \{\overrightarrow{M_{1z}M_{2z}}\} = \gamma_2 - \gamma_1.$$

Por esto la magnitud de las proyecciones del vector sobre los ejes de coordenadas coinciden con las coordenadas de este vector.

La segunda de las fórmulas (23.4) permite calcular, a partir de las coordenadas de los puntos M_1 y M_2, las longitudes de las proyecciones del vector $\overrightarrow{M_1M_2}$ sobre los ejes de coordenadas. A saber,

$$|\overrightarrow{M_{1x}M_{2x}}| = |\alpha_2 - \alpha_1|, \quad |\overrightarrow{M_{1y}M_{2y}}| = |\beta_2 - \beta_1|, \quad |\overrightarrow{M_{1z}M_{2z}}| = |\gamma_2 - \gamma_1|.$$

Longitud de un vector. Deduzcamos la fórmula para calcular la longitud de un vector en el espacio. Es obvio que la longitud $|\overrightarrow{M_1M_2}|$ del vector $\overrightarrow{M_1M_2}$ es igual a la distancia $\rho\ (M_1,\ M_2)$ entre los puntos

M_1, M_2 y equivale también a la longitud de la diagonal de un paralelepípedo rectangular cuyas caras son paralelas a los planos de coordenadas y pasan por los puntos M_1 y M_2 (fig. 25.1). La longitud de cualquier arista del paralelepípedo es igual a la de la proyección del

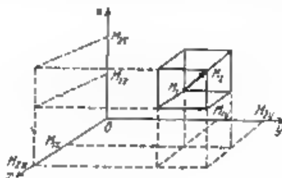


Fig. 25.1.

vector $\overrightarrow{M_1M_2}$ sobre el eje de coordenadas paralelo a la arista. Por ende, haciendo uso del teorema de Pitágoras, obtenemos

$$|\overrightarrow{M_1M_2}| = (|\overrightarrow{M_{1x}M_{2x}}|^2 + |\overrightarrow{M_{1y}M_{2y}}|^2 + |\overrightarrow{M_{1z}M_{2z}}|^2)^{1/2}.$$

Ahora, si los puntos M_1, M_2 están definidos por sus coordenadas $M_1(\alpha_1, \beta_1, \gamma_1)$ y $M_2(\alpha_2, \beta_2, \gamma_2)$, entonces

$$\rho(M_1, M_2) = ((\alpha_2-\alpha_1)^2 + (\beta_2-\beta_1)^2 + (\gamma_2-\gamma_1)^2)^{1/2}. \qquad (25.3)$$

Si el vector $\overrightarrow{M_1M_2}$ está definido por las coordenadas x, y, z respecto de la base $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, entonces

$$|\overrightarrow{M_1M_2}| = (x^2+y^2+z^2)^{1/2}. \qquad (25.4)$$

Una forma análoga la tienen las fórmulas también en el caso de un plano. Si los puntos $M_1(\alpha_1, \beta_1)$, $M_2(\alpha_2, \beta_2)$ o el vector $\overrightarrow{M_1M_2} = (x, y)$ vienen dados mediante sus coordenadas, entonces

$$\rho(M_1, M_2) = ((\alpha_2-\alpha_1)^2 + (\beta_2-\beta_1)^2)^{1/2}, \quad |\overrightarrow{M_1M_2}| = (x^2+y^2)^{1/2}.$$

Ángulo formado por vectores. Examinemos los vectores no nulos $\mathbf{a}$, $\mathbf{b}$ en el espacio. Apliquémoslos al punto O. Designemos con π el plano que pasa por el punto O y contiene ambos vectores. Se llama *ángulo formado por dos vectores* $\mathbf{a}$, $\mathbf{b}$ al ángulo *mínimo* que se obtiene girando alrededor del punto O uno de los vectores en el plano π para que su dirección coincida con la del otro vector. Si al menos uno de los vectores es nulo, el ángulo *no está definido*. Nuestra tarea consistirá en calcular el coseno del ángulo entre dos vectores, a partir de

las coordenadas de dichos vectores. Para el coseno admitamos la designación cos $\{a, b\}$.

Denotaremos mediante A, B los extremos de los vectores a, b en el plano π. Evidentemente, el ángulo formado por los vectores a, b no es otra cosa que el ángulo AOB del triángulo AOB cuyos lados constituyen los vectores a, b y $b-a$ (fig. 25.2).



Fig. 25.2.

Supongamos que los vectores a, b están dados mediante sus coordenadas

$$a = (x_1, y_1, z_1), \qquad b = (x_2, y_2, z_2).$$

En este caso

$$b - a = (x_2 - x_1, y_2 - y_1, z_2 - z_1).$$

Según se sabe de la geometría elemental, el cuadrado de la longitud de un lado del triángulo es igual a la suma de los cuadrados de las longitudes de los otros dos lados menos el producto duplicado de las longitudes de estos lados por el coseno del ángulo entre ellos. Por esta razón

$$|b-a|^2 = |a|^2 + |b|^2 - 2\,|a|\cdot|b| \cos\{a, b\}$$

o bien, al tomar en consideración la fórmula (25.4),

$$(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2 = x_1^2 + y_1^2 + z_1^2 +$$
$$+ x_2^2 + y_2^2 + z_2^2 - 2(x_1^2 + y_1^2 + z_1^2)^{1/2}(x_2^2 + y_2^2 + z_2^2)^{1/2} \cos\{a, b\}.$$

Realizando ciertas transformaciones elementales hallamos

$$\cos\{a, b\} = \frac{x_1x_2 + y_1y_2 + z_1z_2}{(x_1^2+y_1^2+z_1^2)^{1/2}(x_2^2+y_2^2+z_2^2)^{1/2}} \tag{25.5}$$

Los cambios en la fórmula para el caso de un plano son obvios.

División de un segmento en la razón dada. Supongamos que en un espacio se han dado una recta y dos puntos distintos en la misma M_1 y M_2. Elijamos en dicha recta la dirección positiva. En el eje obtenido los puntos M_1 y M_2 determinan un segmento dirigido $\overrightarrow{M_1M_2}$. Sea M un punto cualquiera del eje, distinto de M_2. El número

$$\lambda = \frac{\{\overrightarrow{M_1M}\}}{\{\overrightarrow{MM_2}\}} \tag{25.6}$$

se llama *razón en que el punto M divide el segmento dirigido* $\overrightarrow{M_1M_2}$.

Con el cambio de dirección en el eje, los números $\{\overrightarrow{M_1M}\}$ y $\{\overrightarrow{MM_2}\}$ cambian de signo simultáneamente. Por lo tanto, la razón (25.6) no depende de la dirección positiva elegida en el eje. Luego, con el cambio de la escala de longitudes de los segmentos en el eje, los

números $\{\overrightarrow{M_1M}\}$ y $\{\overrightarrow{MM_2}\}$ se multiplican por el mismo número. Por lo tanto, la razón (25.6) no depende de la unidad elegida para medir longitudes. De aquí se desprende que la razón (25.6) *no depende de cómo se elige en el eje el sistema de coordenadas.*

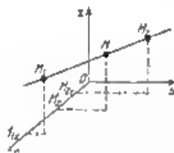


Fig. 25.3.

El problema consiste en calcular las coordenadas del punto M, que divide al segmento $\overrightarrow{M_1M_2}$ en la razón λ, si se conocen las coordenadas de los puntos M_1, M_2 y el número λ, con la particularidad de que $\lambda \neq -1$. Así pues, supongamos que están dados M_1 $(\alpha_1, \beta_1, \gamma_1)$, M_2 $(\alpha_2, \beta_2, \gamma_2)$ y se desconoce M (α, β, γ). Proyectemos estos puntos sobre los ejes de coordenadas, por ejemplo, sobre el eje Ox (fig. 25.3). De los razonamientos de semejanza se ve que el punto M_x divide el segmento dirigido $\overrightarrow{M_{1x}M_{2x}}$ también en la razón λ. Por ello,

$$\lambda = \frac{\{\overrightarrow{M_{1x}M_x}\}}{\{\overrightarrow{M_xM_{2x}}\}}. \qquad (25.7)$$

De conformidad con la fórmula (23.4), $\{\overrightarrow{M_{1x}M_x}\} = \alpha - \alpha_1$, $\{\overrightarrow{M_xM_{2x}}\} = \alpha_2 - \alpha$. Ahora, teniendo en cuenta la expresión (25.7), hallemos que $\alpha = (\alpha_1 + \lambda\alpha_2)/(1 + \lambda)$. De modo análogo se calculan las coordenadas β y γ. Así pues,

$$\alpha = \frac{\alpha_1 + \lambda\alpha_2}{1+\lambda}, \qquad \beta = \frac{\beta_1 + \lambda\beta_2}{1+\lambda}, \qquad \gamma = \frac{\gamma_1 + \lambda\gamma_2}{1+\lambda}.$$

Observamos que $\lambda > 0$, si el punto M se halla dentro del segmento $\overrightarrow{M_1M_2}$; $\lambda < 0$, si el punto M se encuentra fuera del segmento $\overrightarrow{M_1M_2}$, y $\lambda = 0$, si el punto M coincide con M_1. Cuando el punto M se desplaza desde el punto M_1 hasta el punto M_2 (excluyendo el punto M_2), la razón λ toma primero el valor nulo y luego, de la manera sucesiva, todos los valores positivos posibles siempre crecientes. Si el punto M se desplaza desde el punto M_2 en dirección positiva del eje (véase la fig. 25.3), la razón λ tomará primero el valor cero y a continuación valores negativos siempre decrecientes, aproximándose tan cerca como se quiera al valor de $\lambda = -1$, pero quedando en todo momento mayor que dicho valor. Si el punto M se desplaza en dirección negativa desde el punto M_1, la razón λ toma todos los valores negativos posibles en el orden de crecimiento, quedando siempre inferior a $\lambda = -1$.

De esta manera, entre todos los números reales y los puntos de la recta se podría establecer una correspondencia biunívoca, si en la recta hubiera un punto M que divida el segmento $\overrightarrow{M_1M_2}$ en la razón $\lambda = -1$ y si al punto M, coincidente con M_2, se le pudiera poner en correspondencia algún número. Este problema se resuelve comúnmente al completar la recta con un "punto" complementario convencional y los "números", con un "número" complementario convencional. Tal punto se denomina "infinito" y el número, "infinitamente grande".

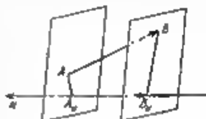


Fig. 25.4.

Proyecciones ortogonales del vector. Sean dados en el espacio cierto eje u y un segmento dirigido $\overrightarrow{AB}$. Tracemos por los puntos A, B unos planos perpendiculares al eje u (fig. 25.4). La intersección de estos planos con el eje determina los puntos A_u, B_u, de los cuales A_u se ubica en el mismo plano con A, mientras que B_u se halla en el mismo plano con B. El segmento dirigido $\overrightarrow{A_uB_u}$ se llama *proyección ortogonal del segmento $\overrightarrow{AB}$ sobre el eje u*. Para designarlo se utiliza el siguiente símbolo:

$$\overrightarrow{A_uB_u} = \text{pr}_u\, \overrightarrow{AB}.$$

Con el eje u fijado, todo vector x del espacio define unívocamente su proyección ortogonal x'. Se puede considerar, por ende, que tenemos una "función" dada

$$x' = \text{pr}_u\, x, \tag{25.8}$$

cuyo "argumento" puede constituir cualquier vector del espacio y el "valor", un vector en el eje u. Demostraremos ahora que dicha función posee las siguientes propiedades:

$$\left.\begin{aligned} \text{pr}_u\,(x+y) &= \text{pr}_u\, x + \text{pr}_u\, y, \\ \text{pr}_u\,(\lambda x) &= \lambda\, \text{pr}_u\, x, \end{aligned}\right\} \tag{25.9}$$

válidas para cualesquiera vectores x e y, como también para todo número λ.

En efecto, fijemos un sistema rectangular cartesiano de coordenadas en el que el eje u coincide con el eje coordenado de abscisas. Supongamos que en este sistema

$$x = (\alpha_1,\ \beta_1,\ \gamma_1),$$
$$y = (\alpha_2,\ \beta_2,\ \gamma_2),$$

entonces

$$x + y = (\alpha_1 + \alpha_2,\ \beta_1 + \beta_2,\ \gamma_1 + \gamma_2),$$
$$\lambda x = (\lambda\alpha_1,\ \lambda\beta_1,\ \lambda\gamma_1).$$

En el sistema de coordenadas elegido la proyección ortogonal del vector sobre el eje u coincide con su proyección coordenada sobre el eje de abscisas. Como se ha indicado antes, la proyección de cualquier vector sobre el eje de abscisas tiene la primera coordenada coincidente con la primera coordenada del propio vector, mientras que las coordenadas restantes son nulas. Por eso

$$\begin{aligned} &\mathrm{pr}_u (x + y) = (\alpha_1 + \alpha_2,\ 0,\ 0),\\ &\mathrm{pr}_u (\lambda x) = (\lambda\alpha_1,\ 0,\ 0),\\ &\mathrm{pr}_u\, x = (\alpha_1,\ 0,\ 0),\\ &\mathrm{pr}_u\, y = (\alpha_2,\ 0,\ 0). \end{aligned} \qquad (25.10)$$

De acuerdo con la regla de adición de vectores y multiplicación de éstos por un número, de las últimas dos igualdades (25.10) concluimos que

$$\mathrm{pr}_u\, x + \mathrm{pr}_u\, y = (\alpha_1 + \alpha_2,\ 0,\ 0),$$
$$\lambda\, \mathrm{pr}_u\, x = (\lambda\alpha_1,\ 0,\ 0).$$

Comparando los segundos miembros de las igualdades obtenidas con los miembros correspondientes de las primeras dos (25.10), nos convencemos de que ambas propiedades (25.9) son justas.

Sean dados en el espacio un plano π y un segmento dirigido $\overrightarrow{AB}$. Al trazar perpendiculares de los puntos A y B sobre el plano π, obtendremos en el plano dado dos puntos A_π y B_π que determinan el segmento dirigido $\overrightarrow{A_\pi B_\pi}$. Este segmento lleva el nombre de *proyección ortogonal del segmento dirigido AB sobre el plano π*. Para designarlo se utiliza la misma notación, es decir,

$$\overrightarrow{A_\pi B_\pi} = \mathrm{pr}_\pi\, \overrightarrow{AB}.$$

Por supuesto, para las proyecciones ortogonales sobre un mismo plano tienen lugar correlaciones análogas a (25.9). Para demostrar esta afirmación, se puede fijar un sistema rectangular cartesiano de coordenadas en el que el plano π sea un plano coordenado y otra vez, hacer uso de las propiedades correspondientes de las proyecciones sobre un plano de coordenadas.

Hemos considerado las proyecciones ortogonales de los vectores en el espacio. Indudablemente, la analogía completa tiene lugar también para los vectores en un plano.

Ejercicios.

1. Dos vectores no nulos están dados mediante sus coordenadas cartesianas. ¿En qué casos serán perpendiculares entre sí?

2. Hállense las coordenadas del centro de gravedad de tres puntos materiales, si se conocen sus coordenadas cartesianas y sus masas.

3. Hállese el área de un triángulo, si se conocen las coordenadas cartesianas de sus tres vértices.

4. En un espacio están dados los vectores no nulos x, a, b, c, con la particularidad de que a, b, c son perpendiculares dos a dos. Demuéstrese que

$$\cos^2(x, a) + \cos^2(x, b) + \cos^2(x, c) = 1.$$

5. Designemos con π un plano coordenado cualquiera y con u, un eje coordenado cualquiera en el plano π. Demuéstrese que para cualquier vector a se verifica

$$\mathrm{pr}_u(\mathrm{pr}_\pi a) = \mathrm{pr}_u a.$$

§ 26. Producto escalar

El uso de los segmentos dirigidos para expresar fuerzas y desplazamientos conduce a un concepto muy importante de producto escalar de vectores.

Del curso de la física sabemos que si un vector a representa una fuerza tal, que el punto de su aplicación se desplaza desde el origen del vector b al extremo del mismo, el trabajo ω de esta fuerza se determinará por la igualdad

$$\omega = |a|\,|b| \cos(a, b). \tag{26.1}$$

El segundo miembro de esta igualdad se llama *producto escalar de los vectores a, b*. Para su designación se ha aceptado el símbolo (a, b). De este modo,

$$(a, b) = |a|\,|b| \cos(a, b). \tag{26.2}$$

Hablando en rigor, la definición aducida de producto escalar se refiere sólo a los vectores no nulos a, b, puesto que solamente para vectores de tal tipo está definido el ángulo. Sin embargo, tomando en consideración la preimagen del producto escalar, es fácil comprender cuál ha de ser la definición adicional en el caso en que siquiera uno de los vectores sea igual a cero. Si bien una fuerza, bien un desplazamiento, se da mediante un vector nulo, el trabajo que se cumple es nulo. Por esta razón consideraremos que $(a, b) = 0$, siempre que al menos uno de los vectores a, b sea igual a cero.

De la fórmula (26.2) provienen ciertas propiedades geométricas del producto escalar. Por ejemplo, el ángulo, formado por dos vectores no nulos, será agudo (obtuso) cuando, y sólo cuando, el producto escalar de estos vectores es positivo (negativo).

Si el ángulo formado por los vectores es recto o al menos uno de los vectores es nulo, el producto escalar será igual a cero. Los vectores de este tipo se llamarán *ortogonales*.

Se denominarán vectores *ortonormalizados* aquellos vectores ortogonales cuya longitud es igual a la unidad. En particular, los vectores básicos i, j, k de un sistema rectangular cartesiano de coordenadas son ortonormalizados. De la fórmula (26.2) se deduce que

$$\begin{aligned} &(i, i)=1, \quad (i, j)=0, \quad (i, k)=0, \\ &(j, i)=0, \quad (j, j)=1, \quad (j, k)=0, \\ &(k, i)=0, \quad (k, j)=0, \quad (k, k)=1. \end{aligned} \qquad (26.3)$$

Examinemos los vectores no nulos a, b. Por el vector a tracemos el eje u atribuyéndole una dirección tal que la magnitud del vector a sea positiva. Será evidente que

$$(\mathrm{pr}_u b)=|b| \cos (a, b).$$

La proyección del vector b sobre un eje construido de este modo, la llamaremos *proyección del vector b sobre el vector a* y la indicaremos mediante el símbolo $\mathrm{pr}_a b$. Naturalmente, la proyección de un vector sobre otro conserva las propiedades (25.9). En las designaciones nuevas

$$(a, b)=|a|(\mathrm{pr}_a b)=|b|(\mathrm{pr}_b a). \qquad (26.4)$$

Estas fórmulas permiten establecer propiedades algebraicas muy importantes del producto escalar. A saber, para cualesquiera vectores a, b, c y todo número real α se verifican las correlaciones:

$$\begin{aligned} &1)\ (a, b)=(b, a), \\ &2)\ (\alpha a, b)=\alpha(a, b), \\ &3)\ (a+b, c)=(a, c)+(b, c), \\ &4)\ (a, a)>0 \text{ cuando } a \neq 0; \ (0, 0)=0. \end{aligned} \qquad (26.5)$$

Hemos de notar que las correlaciones (26.5) se cumplen, a ciencia cierta, si por lo menos uno de los vectores es nulo. En el caso general la validez de las propiedades 1,4 se desprende directamente de la fórmula (26.2). En cuanto a las propiedades 2,3, haremos uso de las fórmulas (26.4) y las propiedades de proyecciones, para establecerlas. Tenemos

$$(\alpha a, b)=|b|(\mathrm{pr}_b(\alpha a))=|b|(\alpha \cdot \mathrm{pr}_b a)=\alpha|b|(\mathrm{pr}_b a)=\alpha(a, b),$$

$$(a+b, c)=|c|(\mathrm{pr}_c(a+b))=|c|(\mathrm{pr}_c a+\mathrm{pr}_c b)=$$
$$=|c|(\mathrm{pr}_c a)+|c|(\mathrm{pr}_c b)=(a, c)+(b, c).$$

Las propiedades 2,3 sólo están ligadas con el primer factor del producto escalar. Las propiedades análogas tienen lugar también respecto del segundo factor. En efecto,

$$(a, \alpha b)=(\alpha b, a)=\alpha(b, a)=\alpha(a, b),$$
$$(a, b+c)=(b+c, a)=(b, a)+(c, a)=(a, b)+(a, c).$$

Además, en virtud de la igualdad $a - b = a + (-1)\,b$, serán lícitas también las siguientes correlaciones:

$$(a - b,\ c) = (a,\ c) - (b,\ c),$$
$$(a,\ b - c) = (a,\ b) - (a,\ c)$$

puesto que

$$(a - b,\ c) = (a + (-1)\,b,\ c) = (a,\ c) + ((-1)\,b,\ c) =$$
$$= (a,\ c) + (-1)\,(b,\ c) = (a,\ c) - (b,\ c).$$

TEOREMA 26.1. *Si dos vectores a, b están dados mediante sus coordenadas rectangulares cartesianas, entonces el producto escalar de estos vectores es igual a la suma de los productos, realizados dos a dos, de las coordenadas correspondientes.*

DEMOSTRACIÓN. Supongamos, para concretar, que los vectores están dados en un espacio, es decir, $a = (x_1,\ y_1,\ z_1)$, $b = (x_2,\ y_2,\ z_2)$. Puesto que

$$a = x_1 i + y_1 j + z_1 k,$$
$$b = x_2 i + y_2 j + z_2 k,$$

entonces, al efectuar las transformaciones algebraicas del producto escalar, obtenemos

$$(a,\ b) = x_1x_2\,(i,\ i) + x_1y_2\,(i,\ j) + x_1z_2\,(i,\ k) +$$
$$+ y_1x_2\,(j,\ i) + y_1y_2\,(j,\ j) + y_1z_2\,(j,\ k) +$$
$$+ z_1x_2\,(k,\ i) + z_1y_2\,(k,\ j) + z_1z_2\,(k,\ k).$$

Ahora, de acuerdo con (26.3) se tiene

$$(a,\ b) = x_1x_2 + y_1y_2 + z_1z_2, \qquad (26.6)$$

y el teorema queda demostrado.

La fórmula (26.6) permite escribir las expresiones, recién obtenidas, (25.4), (25.5) para la longitud del vector y el ángulo formado por los vectores en términos de un producto escalar. A saber,

$$|a| = (a,\ a)^{1/2}, \qquad (26.7)$$
$$\cos(a,\ b) = \frac{(a,\ b)}{|a|\cdot|b|}.$$

Puede parecer que estas fórmulas son triviales, puesto que se desprenden directamente de (26.2) sin referencia alguna a las fórmulas (25.4), (25.5). No obstante, sin darnos prisa en hacer tal deducción, veamos una circunstancia de importancia particular.

Ha de ser notado que toda nuestra investigación se desarrollaba, de hecho, en tres etapas. En primer lugar, partiendo de la fórmula (26.2), hemos demostrado la validez de las propiedades (26.5). A continuación, basándonos solamente en estas propiedades y en el carác-

tor ortonormal de los vectores básicos del sistema de coordenadas, deducimos la fórmula (26.6). Y, por fin, haciendo uso de las fórmulas (25.4), (25.5), que se han obtenido sin referencias al concepto de producto escalar de los vectores, llegamos a las fórmulas (26.7).

Al partir de lo dicho, podríamos ahora introducir el producto escalar no por definición de su forma explícita, sino *de un modo axiomático*, como cierta función numérica definida para todo par de vectores, exigiendo en tal caso que las propiedades (26.5) se cumplan sin falta. Entonces, para cualesquiera sistemas de coordenadas, donde los vectores básicos están ortonormalizados en el sentido de producto escalar axiomático, seguirá siendo válida la correlación (26.6). Por consiguiente, teniendo presente un modelo del sistema rectangular cartesiano de coordenadas, podríamos considerar axiomáticamente que las longitudes de los vectores y los ángulos entre ellos se calculan según las fórmulas (26.7). Desde luego, tendríamos que convencernos, en tal caso, de que las longitudes y los ángulos que se han introducido de modo semejante, poseen las propiedades necesarias.

Ejercicios.

1. Están dados dos vectores a y b. ¿Bajo qué condiciones impuestas al número α los vectores a y $b + \alpha a$ son ortogonales? ¿Cuál es la interpretación geométrica de este problema?

2. El vector a se ha dado en el espacio V_3 mediante sus coordenadas cartesianas. Hállense dos vectores linealmente independientes ortogonales al vector a.

3. Los vectores linealmente independientes a, b están dados en el espacio V_3 mediante sus coordenadas cartesianas. Hállese el vector no nulo que sea ortogonal respecto de ambos vectores.

4. ¿Qué representa en sí el lugar geométrico de vectores ortogonales a un vector dado?

§ 27. Espacio euclídeo

Los espacios lineales abstractos estudiados anteriormente son, en cierto sentido, más pobres de conceptos y propiedades en comparación con el espacio de segmentos dirigidos. Esta pobreza se debe, en primer lugar, a que en los primeros no se han reflejado los hechos más importantes relacionados con la medición de longitudes, de ángulos, de volúmenes, etc. Los conceptos métricos pueden ser extendidos a los espacios lineales abstractos por medio de varios procedimientos. No obstante, un método más eficaz para posibilitar la realización de mediciones consiste en la introducción *axiomática* de un producto escalar de vectores. Comenzaremos nuestras investigaciones con los espacios lineales reales.

Un espacio lineal real E se llama *euclídeo*, si a todo par de vectores x, y de E se le ha puesto en correspondencia un número real (x, y), llamado *producto escalar*, considerándose cumplidos los siguientes

axiomas:

$$\begin{aligned} &1)\ (x, y) = (y, x),\\ &2)\ (\lambda x, y) = \lambda (x, y),\\ &3)\ (x+y, z) = (x, z) + (y, z),\\ &4)\ (x, x) > 0 \text{ para } x \neq 0;\ (0, 0) = 0 \end{aligned} \qquad (27.1)$$

para los vectores arbitrarios x, y, z de E y el número real arbitrario λ.

Como ya sabemos de estos axiomas se desprende que con el producto escalar se pueden realizar transformaciones algebraicas formales, es decir,

$$\left(\sum_{i=1}^{s} \alpha_i x_i,\ \sum_{j=1}^{t} \beta_j y_j\right) = \sum_{i=1}^{s} \sum_{j=1}^{t} \alpha_i \beta_j (x_i, y_j)$$

para cualesquiera vectores x_i, y_j, los números α_i, β_j y para todo número s ó t de sumandos.

Todo subespacio lineal L del espacio euclídeo E se convierte por sí mismo en un espacio euclídeo, si se conserva en L el producto escalar que se ha introducido en E.

Es fácil indicar el método general de introducción del producto escalar en un espacio real arbitrario K. Sea $e_1, e_2, \ldots, e_n$ una base de este espacio. Elijamos dos vectores arbitrarios x, y de K y supongamos que

$$x = \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n,$$
$$y = \eta_1 e_1 + \eta_2 e_2 + \ldots + \eta_n e_n.$$

El producto escalar de vectores puede ser introducido ahora del modo siguiente, por ejemplo:

$$(x, y) = \xi_1 \eta_1 + \xi_2 \eta_2 + \ldots + \xi_n \eta_n. \qquad (27.2)$$

El cumplimiento de todos los axiomas se comprueba sin dificultad alguna. Por consiguiente, el espacio lineal K provisto del producto escalar (27.2) es euclídeo.

Observemos que el producto escalar puede ser introducido en el espacio K con ayuda de otros métodos. Por ejemplo, en dicho espacio la siguiente expresión será también un producto escalar:

$$(x, y) = \alpha_1 \xi_1 \eta_1 + \alpha_2 \xi_2 \eta_2 + \ldots + \alpha_n \xi_n \eta_n$$

para cualesquiera números positivos fijados $\alpha_1, \alpha_2, \ldots, \alpha_n$. La multiformidad semejante no debe disturbarnos. En efecto, no hay nada asombroso en que las longitudes puedan medirse en metros y en pulgadas, los ángulos en grados y radianes, etc. Precisamente esta multiformidad permite emplear al máximo las propiedades de los espacios concretos, cuando en éstos se introduce el producto escalar.

Al introducir el producto escalar en los espacios de segmentos dirigidos, tuvimos que definirlo por separado, cuando por lo menos

uno de los segmentos era igual a cero. En aquel caso se suponía que el producto escalar era nulo. Ahora el hecho dado pasa a ser una propiedad que proviene de los axiomas (27.1). Si x es un vector arbitrario de E, entonces

$$(0, x) = (0x, x) = 0\,(x, x) = 0.$$

Por supuesto, en virtud del primer axioma de (27.1), $(x, 0) = 0$.

Un vector x de un espacio euclídeo se denomina *normalizado*, si $(x, x) = 1$. Todo vector *no nulo* y puede ser normalizado, si lo multiplicamos por cierto número λ. En efecto, por hipótesis

$$(\lambda y, \lambda y) = \lambda^2 (y, y) = 1,$$

y por eso, a título del factor de normalización, podemos tomar

$$\lambda = (y, y)^{-1/2}.$$

Un sistema de vectores se llama *normalizado*, si están normalizados todos sus vectores. Según se deduce de lo dicho, todo sistema de vectores no nulos puede ser normalizado.

Una de las propiedades más importantes del producto escalar la enuncia el siguiente

TEOREMA 27.1 (*desigualdad de Cauchy—Buniakovski*). *Para cualesquiera dos vectores x, y de un espacio euclídeo es válida la desigualdad*

$$(x, y)^2 \leqslant (x, x)\,(y, y).$$

DEMOSTRACIÓN. El teorema tiene lugar, a ciencia cierta, si $y = 0$, por lo cual convengamos en considerar que $y \neq 0$. Examinemos un vector $x - \lambda y$, donde λ es un número real arbitrario. Tenemos

$$(x - \lambda y,\ x - \lambda y) = (x, x) - 2\lambda\,(x, y) + \lambda^2\,(y, y).$$

En el primer miembro de la igualdad figura un producto escalar de vectores iguales. Por esta razón el trinomio de segundo grado en el segundo miembro es no negativo, cualquiera que sea λ, en particular, para

$$\lambda = \frac{(x, y)}{(y, y)}. \qquad (27.3)$$

De este modo,

$$(x, x) - 2\frac{(x, y)}{(y, y)}\,(x, y) + \frac{(x, y)^2}{(y, y)^2}\,(y, y) = (x, x) - \frac{(x, y)^2}{(y, y)} \geqslant 0,$$

de donde se hace obvia la afirmación del teorema.

Por analogía con los espacios de segmentos dirigidos, llamemos *colineales* dos vectores x, y de cualquier espacio lineal, si o bien $x = \lambda y$ o bien $y = \mu x$ para ciertos números λ, μ. En virtud de la igualdad $0 = 0x$ concluimos que los dos vectores son colineales, a ciencia cierta, si por lo menos uno de ellos es nulo. Para la comprobación del carácter colineal de los vectores resulta muy cómoda la desigualdad de Cauchy—Buniakovski. En particular, es válido el

TEOREMA 27.2. *La desigualdad de Cauchy—Buniakovski se convierte en una igualdad si, y sólo si, los vectores x, y son colineales.*

DEMOSTRACIÓN. Supongamos que los vectores x, y son colineales y que, para concretar, $x = \lambda y$. Hallamos

$$(x, y)^2 = (\lambda y, y)^2 = \lambda^2 (y, y)^2,$$
$$(x, x)(y, y) = (\lambda y, \lambda y)(y, y) = \lambda^2 (y, y)^2.$$

La comparación de estas igualdades muestra que la suficiencia de la afirmación del teorema tiene lugar.

Supongamos ahora que para ciertos vectores x, y se verifica la igualdad:

$$(x, y)^2 = (x, x)(y, y). \tag{27.4}$$

Si $y = 0$, los vectores son colineales. En cambio, si $y \neq 0$, entonces, al tomar λ de acuerdo con (27.3) y teniendo presente (27.4), obtenemos

$$(x - \lambda y, x - \lambda y) = 0.$$

En vista del último axioma en (27.1), esto significa que $x - \lambda y = 0$, o bien $x = \lambda y$, es decir, los vectores x, y son colineales. La necesidad de la afirmación del teorema también tiene lugar.

A título de ejemplo consideremos el espacio R_n. Se le puede transformar en espacio euclídeo, si para los vectores

$$x = (\alpha_1, \alpha_2, \ldots, \alpha_n),$$
$$y = (\beta_1, \beta_2, \ldots, \beta_n)$$

el producto escalar se introduce de la manera siguiente:

$$(x, y) = \sum_{i=1}^{n} \alpha_i \beta_i. \tag{27.5}$$

Es evidente que los axiomas (27.1) se cumplen. La desigualdad de Cauchy—Buniakovski significa, en el caso dado, que

$$\left(\sum_{i=1}^{n} \alpha_i \beta_i\right)^2 \leqslant \left(\sum_{i=1}^{n} \alpha_i^2\right)\left(\sum_{i=1}^{n} \beta_i^2\right) \tag{27.6}$$

para cualesquiera números reales α_i, β_i.

Ejercicios.

1. ¿Cómo se debe introducir el producto escalar en un espacio de polinomios de una variable con coeficientes reales?
2. ¿Se hará euclídeo el espacio R_n, si el producto escalar se introduce en él de la manera siguiente:

$$(x, y) = \sum_{i=1}^{n} |\alpha_i| \, |\beta_i| ?$$

3. ¿Cuál es el sentido geométrico de la desigualdad de Cauchy—Buniakovski en el espacio de segmentos dirigidos?
4. Demuéstrese que $x = y$ cuando, y sólo cuando, $(x, d) = (y, d)$ para todo vector d.

§ 28. Ortogonalidad

La relación más importante entre los vectores de un espacio euclídeo es la ortogonalidad.

Por definición, los vectores x, y se denominan *ortogonales* si $(x, y) = 0$. En virtud del primer axioma de (27.1), la relación de ortogonalidad de dos vectores es simétrica. En el espacio de segmentos dirigidos la noción de ortogonalidad coincide, en lo principal, con la de perpendicularidad. Por esta razón, la ortogonalidad puede considerarse como una generalización de la noción de perpendicularidad para los espacios euclídeos abstractos.

Un sistema de vectores de un espacio euclídeo se llama *ortogonal* si o bien consiste en un solo vector o bien sus vectores son ortogonales dos a dos. Si un sistema ortogonal consta de vectores no nulos, se lo puede normalizar. Un sistema ortogonal normalizado se denomina *ortonormalizado*.

El interés hacia los sistemas ortogonales y ortonormalizados se debe a las ventajas que ofrecen en la investigación de espacios euclídeos.

Así por ejemplo, cualquier sistema ortogonal de vectores no nulos y, por supuesto, un sistema ortonormalizado, es linealmente independiente. En efecto, supongamos que el sistema $x_1, x_2, \ldots, x_k$ es ortogonal y que $x_i \neq 0$ para todo i. Esto significa que $(x_i, x_j) = 0$ para $i \neq j$, pero $(x_i, x_j) \neq 0$ cuando $i = j$. Escribamos la ecuación

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k = 0.$$

Al multiplicarla escalarmente por cualquiera de los vectores x_i hallamos

$$\alpha_1 (x_1, x_i) + \alpha_2 (x_2, x_i) + \ldots + \alpha_k (x_k, x_i) = 0.$$

Por consiguiente,

$$\alpha_i (x_i, x_i) = 0 \qquad (28.1)$$

y, naturalmente, $\alpha_i = 0$. De este modo, el sistema de los vectores $x_1, x_2, \ldots, x_k$ es linealmente independiente.

De la igualdad (28.1) obtenemos, en particular, que si la suma de los vectores ortogonales dos a dos es igual a cero, todos los vectores son nulos.

Una variedad de corolarios útiles se deduce de la suposición que cierto sistema ortonormalizado $e_1, e_2, \ldots, e_n$ puede formar una base del espacio euclídeo E. En este caso todo vector x de E ha de representarse de un modo único, en forma de una combinación lineal

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n.$$

Pero al multiplicar escalarmente la igualdad dada por e_i obtenemos la expresión explícita para los coeficientes de la descomposición

según la base. A saber,

$$\alpha_i = (x, e_i). \qquad (28.2)$$

Si para otro vector y tiene lugar la descomposición

$$y = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n,$$

entonces, al realizar ciertas transformaciones sencillas, hallamos que

$$(x, y) = \alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n. \qquad (28.3)$$

En particular,

$$(x, x) = \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2. \qquad (28.4)$$

Antes de continuar unas investigaciones semejantes, aclaremos si existe o no una base que se componga de vectores ortonormalizados.

Una base cuyos vectores formen un sistema ortonormalizado, se denomina *ortonormalizada*. La existencia de tal base en el espacio euclídeo la demuestra el

TEOREMA 28.1. *En todo espacio euclídeo E de dimensión finita existe una base ortonormalizada.*

DEMOSTRACIÓN. Sea $\dim E = n$. Un sistema ortonormalizado es linealmente independiente y, por ende, no puede contener más que n vectores. Supongamos que el sistema $e_1, e_2, \ldots, e_s$ contiene el número máximo de vectores ortonormalizados. Esto significa que en el espacio E no hay ni un solo vector no nulo que sea ortogonal a todos los vectores $e_1, e_2, \ldots, e_s$. Si existe un vector ortogonal a todos estos vectores, el mismo debe ser nulo.

Elijamos un vector arbitrario x de E. Si el sistema ortonormalizado $e_1, e_2, \ldots, e_s$ fuera una base, el vector x tendría que coincidir con el vector y, donde

$$y = (x, e_1)\, e_1 + (x, e_2)\, e_2 + \ldots + (x, e_s)\, e_s.$$

Examinemos, por ello, un vector $x - y$. Tenemos

$$(x - y, e_i) = \Big(x - \sum_{p=1}^{s} (x, e_p)\, e_p, e_i\Big) =$$

$$= (x, e_i) - \sum_{p=1}^{s} (x, e_p)(e_p, e_i) = (x, e_i) - (x, e_i) = 0.$$

El vector $x - y$ resulta ser ortogonal a todos los vectores $e_1, e_2, \ldots, e_s$. Por consiguiente, $x - y = 0$ ó $x = y$.

Así pues, un sistema linealmente independiente $e_1, e_2, \ldots, e_s$ posee la propiedad que, en términos de sus vectores, se expresa linealmente todo vector del espacio E, es decir el sistema forma una base.

COROLARIO. *Cualquier sistema ortonormalizado de vectores $e_1, e_2, \ldots, e_k$ puede ser complementado hasta obtener una base ortonormalizada.*

Efectivamente, de los sistemas ortonormalizados que contienen un sistema dado tomemos el que cuenta con el máximo número de vectores. Supongamos que este sistema es $e_1, \ldots, e_k, e_{k+1}, \ldots, e_r$. Repitiendo textualmente la demostración del teorema 28.1, establecemos que el sistema nuevo es una base.

Además de los vectores ortogonales en el espacio euclídeo, consideraremos también *conjuntos ortogonales de vectores*. Dos conjuntos F y G de vectores en el espacio euclídeo E se denominan ortogonales, si todo vector de F es ortogonal a todo vector de G. La ortogonalidad de F y G se designa con el símbolo $F \perp G$.

Por supuesto, un conjunto puede componerse también de un solo vector. Si cierto vector de un conjunto es ortogonal a todo el conjunto, es, en particular, ortogonal a sí mismo. Por consiguiente, sólo puede ser nulo.

LEMA 29.1. *Para que el vector x sea ortogonal al subespacio L, es necesario y suficiente que sea ortogonal respecto de todos los vectores de una base cualquiera del subespacio L.*

DEMOSTRACIÓN. Fijemos la base $y_1, y_2, \ldots, y_k$ del subespacio L. Si $x \perp L$, entonces x es ortogonal a todos los vectores de L y, en particular, a los vectores $y_1, y_2, \ldots, y_k$. Sea, ahora, $(x, y_i) = 0$ para todo i. Elijamos un vector arbitrario z de L y descompongámoslo según los vectores de la base. Si

$$z = \alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_k y_k$$

para ciertos números $\alpha_1, \alpha_2, \ldots, \alpha_k$, entonces

$$(x, z) = (x, \alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_k y_k) =$$
$$= \alpha_1 (x, y_1) + \alpha_2 (x, y_2) + \ldots + \alpha_k (x, y_k) = 0.$$

Esto significa que $x \perp L$.

COROLARIO. *Para que dos subespacios sean ortogonales, es necesario y suficiente que todo vector de cualquier base de un subespacio sea ortogonal a todos los vectores de cualquier base de otro subespacio.*

La suma K de los subespacios lineales $L_1, L_2, \ldots, L_m$ se llama *ortogonal*, si los subespacios son ortogonales dos a dos. Para designar la suma ortogonal se empleará la siguiente notación:

$$K = L_1 \oplus L_2 \oplus \ldots \oplus L_m.$$

LEMA 29.2. *Una suma ortogonal de los subespacios no nulos es siempre una suma directa.*

DEMOSTRACIÓN. Elijamos una base ortonormalizada en cada subespacio y consideremos un sistema de vectores que representa en sí la reunión de bases de todos los subespacios. Está claro que todo vector de la suma ortogonal se expresa linealmente en términos de los vectores del sistema construido. Mas, este sistema es linealmente independiente, ya que consiste en los vectores no nulos ortogonales dos a dos. Ahora la afirmación del lema se deduce del teorema 20.1.

Supongamos que el espacio euclídeo K figura en forma de una suma ortogonal de sus subespacios $L_1, L_2, \ldots, L_m$, caso en el que la totalidad de todos estos subespacios puede considerarse como una base ortogonal generalizada. En particular, si, para cualesquiera dos vectores x, y de K, escribimos sus descomposiciones según los subespacios $L_1, L_2, \ldots, L_m$, es decir, si representamos los vectores citados en la forma

$$x = x_1 + x_2 + \ldots + x_m,$$
$$y = y_1 + y_2 + \ldots + y_m,$$

donde $x_i, y_i \in L_i$, se obtiene con facilidad que

$$(x, y) = (x_1, y_1) + (x_2, y_2) + \ldots + (x_m, y_m). \qquad (28.5)$$

La fórmula obtenida es análoga a la (28.3).

Examinemos un conjunto no vacío arbitrario de vectores F del espacio euclídeo E. La totalidad de todos los vectores, ortogonales al conjunto F, se llama *complemento ortogonal* del conjunto F y se indica con $F^\perp$. El complemento ortogonal es un subespacio. En efecto, si los vectores $x, y \in F^\perp$, entonces $x, y \perp F$. Pero, en este caso, $\alpha x + \beta y \perp F$ para cualesquiera números α, β, es decir, $\alpha x + \beta y \in F^\perp$.

TEOREMA 28.2. *El espacio euclídeo E es la suma ortogonal de cualquier subespacio lineal suyo L y su complemento ortogonal $L^\perp$, es decir,*

$$E = L \oplus L^\perp.$$

DEMOSTRACIÓN. Sea $\dim L = s$, $\dim L^\perp = m$. Elijamos una base ortonormalizada $e_1, \ldots, e_s$ del subespacio L y una base ortonormalizada $r_1, \ldots, r_m$ del subespacio $L^\perp$. El sistema de vectores $e_1, \ldots, e_s, r_1, \ldots, r_m$ es ortonormalizado y, por lo tanto, linealmente independiente.

Si este sistema no es la base de E, se le puede complementar hasta que se obtenga la base ortonormalizada de E. Sea e uno de los vectores complementarios. Es ortogonal a los vectores $e_1, \ldots, e_s$, por lo cual $e \perp L$, es decir, $e \in L^\perp$. Pero, por otra parte, el vector e es ortogonal a los vectores $r_1, \ldots, r_m$ y, por ende, $e \perp L^\perp$. Así pues, el vector e simultáneamente pertenece a $L^\perp$ y es ortogonal a $L^\perp$. Por consiguiente, $e = 0$, lo que demuestra la afirmación del teorema.

La descomposición de un espacio en una suma ortogonal de sus subespacios permite realizar eficazmente muchas investigaciones. Ilustrémoslo con el ejemplo siguiente.

Elijamos un espacio euclídeo E y examinemos en él un sistema fijado de vectores $x_1, x_2, \ldots, x_k$. Si el rango de este sistema es igual a la dimensión de E, entonces, evidentemente, el único vector de E, ortogonal a todos los vectores del sistema dado, será el vector nulo. Tiene lugar también el lema inverso:

LEMA 28.3. *Si en el espacio euclídeo E se ha dado cierto sistema de vectores $x_1, x_2, \ldots, x_k$ y el el único vector de E, ortogonal a estos vecto-*

res, es nulo, entonces el rango del sistema equivale a la dimensión de E.

DEMOSTRACIÓN. Designemos con L la cápsula lineal del sistema de vectores $x_1, x_2, \ldots, x_k$. Todo vector, ortogonal a los vectores dados, es ortogonal a L, es decir, pertenece al complemento ortogonal de $L^\perp$. De acuerdo con la hipótesis del lema, el subespacio $L^\perp$ sólo consta de un vector nulo. Puesto que $E = L \oplus L^\perp$, de aquí se deduce que la dimensión de L coincide con la de E. Mas, la dimensión de L es igual al rango del sistema de vectores $x_1, x_2, \ldots, x_k$. El lema queda demostrado.

Ejercicios.

1. Demuéstrese que si el producto escalar de cualesquiera dos vectores de un espacio euclídeo se expresa mediante la igualdad (28.3), la base, respecto de la cual se han tomado las coordenadas, es ortonormalizada.
2. Demuéstrese que si el producto escalar de cualquier vector de un espacio euclídeo por sí mismo se expresa mediante la igualdad (28.4), la base, respecto de la cual se han tomado las coordenadas, es ortonormalizada.
3. Demuéstrese que si dos conjuntos compuestos por un número finito de vectores son ortogonales, serán ortogonales también las cápsulas lineales construidas en estos conjuntos.
4. Demuéstrese que la intersección de dos subespacios ortogonales sólo consta de un vector nulo.
5. Demuéstrese que si un espacio euclídeo es la suma directa de sus subespacios y para cualesquiera dos vectores tiene lugar la igualdad (28.5), los subespacios son ortogonales dos a dos.
6. Demuéstrese que para cualesquiera subespacios L, M de un espacio euclídeo E se verifican las correlaciones

$$\dim L + \dim L^\perp = \dim E,$$
$$(L^\perp)^\perp = L,$$
$$(L + M)^\perp = L^\perp \cap M^\perp,$$
$$(L \cap M)^\perp = L^\perp + M^\perp.$$

§ 29. Longitudes, ángulos, distancias

Haremos extender, ahora, las nociones de longitud, ángulo y distancia a los elementos del espacio euclídeo. Realizándolo partiremos de la analogía con los espacios de segmentos dirigidos.

Se llama *longitud* $|x|$ *del vector* x en el espacio euclídeo E la magnitud

$$|x| = +(x, x)^{1/2}.$$

Todo vector tiene su longitud. De acuerdo con el último axioma (27.1), ésta es positiva para los vectores no nulos y es igual a cero, para el vector nulo. Luego, la igualdad

$$|\lambda x| = (\lambda x, \lambda x)^{1/2} = (\lambda^2 (x, x))^{1/2} = |\lambda|\,|x|$$

demuestra la posibilidad de sacar la magnitud absoluta del factor numérico λ fuera del signo de la longitud del vector. Como se ha indicado anteriormente, el vector no nulo puede ser normalizado, es decir, puede ser multiplicado por un número tal que la longitud del vector resultante se haga igual a la unidad.

Se llama *ángulo* $\{x, y\}$ *formado por los vectores no nulos* x, y del espacio euclídeo E aquel que se define por las correlaciones

$$\cos\{x, y\} = \frac{(x, y)}{|x|\,|y|}, \quad 0 \leqslant \{x, y\} \leqslant \pi.$$

Si entre los vectores x, y existe al menos uno no nulo, el ángulo formado por tales vectores se considera *indeterminado*.

La desigualdad de Cauchy—Buniakovski permite afirmar que la expresión, llamada coseno del ángulo entre vectores, no es superior, en módulo, a la unidad. Por ello el ángulo formado por cualesquiera vectores no nulos está siempre definido y, además, unívocamente. Este ángulo no se altera cuando los vectores se multiplican por cualesquiera números positivos y, conforme al teorema 27.2, es igual a 0 ó π si, y sólo si, los vectores no nulos son colineales. Todo esto concuerda plenamente con la noción de ángulo formado por segmentos dirigidos.

Elijamos dos vectores no nulos x, y. Teniendo presente la analogía con los segmentos dirigidos, consideraremos x, y como dos de los lados de cierto triángulo. Como tercer lado del triángulo resulta natural tomar el vector $x - y$. Al hacer uso de la definición de longitud de un vector y de ángulo entre los vectores, hallamos

$$\begin{aligned} |x - y|^2 &= (x - y, x - y) = (x, x) - 2(x, y) + (y, y) = \\ &= |x|^2 + |y|^2 - 2\,|x|\,|y|\cos\{x, y\}. \end{aligned} \qquad (29.1)$$

Hemos probado, pues, que en el espacio euclídeo el cuadrado de la longitud de cualquier lado de un triángulo es igual a la suma de cuadrados de las longitudes de sus otros dos lados menos el producto duplicado de las longitudes de estos lados por el coseno del ángulo entre ellos.

Si el triángulo es rectángulo, es decir, si el ángulo formado por los vectores x, y es recto, entonces, evidentemente,

$$|x - y|^2 = |x|^2 + |y|^2. \qquad (29.2)$$

Esto no es otra cosa que la expresión formal del teorema conocido de Pitágoras.

Consideremos de nuevo un triángulo rectángulo. Puesto que el coseno del ángulo formado por los vectores no sobrepasa en módulo la unidad, entonces, de (29.1) proviene que

$$|x - y|^2 \leqslant (|x| + |y|)^2,$$
$$|x - y|^2 \geqslant (|x| - |y|)^2,$$

o bien

$$|x-y| \leqslant |x| + |y|, \qquad (29.3)$$

$$|x-y| \geqslant ||x| - |y||.$$

De este modo, en el espacio euclídeo la longitud del lado de un triángulo no es superior a la suma de longitudes de otros dos lados, pero no es inferior a la diferencia entre sus longitudes.

Se llama *distancia* $\rho(x, y)$ *entre los vectores* x, y de un espacio euclídeo la magnitud

$$\rho(x, y) = |x-y|. \qquad (29.4)$$

Esta magnitud satisface tres propiedades naturales de la distancia entre los vectores (en la interpretación puntual) en los espacios de segmentos dirigidos. A saber, para cualesquiera vectores x, y, z de un espacio euclídeo

1) $\rho(x, y) = \rho(y, x)$,

2) $\rho(x, y) > 0$, si $x \neq y$; $\rho(x, y) = 0$, si $x = y$, (29.5)

3) $\rho(x, y) \leqslant \rho(x, z) + \rho(z, y)$.

Las primeras dos propiedades son obvias. La última propiedad no es otra cosa que la generalización de la conocida desigualdad triangular. La validez de la propiedad 3 se desprende de la primera desigualdad (29.3), si x e y se sustituyen por $x - z$ e $y - z$, respectivamente.

Se llama *distancia* $\rho(A, B)$ *entre los conjuntos* A, B *de vectores* de un mismo espacio la magnitud

$$\rho(A, B) = \inf_{x \in A,\, y \in B} \rho(x, y).$$

En conclusión hemos de notar la siguiente circunstancia. Supongamos que en el espacio euclídeo E se ha fijado una base ortonormalizada $e_1, e_2, \ldots, e_n$. Para cualesquiera dos vectores x, y, dados mediante sus coordenadas

$$x = (\alpha_1, \alpha_2, \ldots, \alpha_n), \qquad y = (\beta_1, \beta_2, \ldots, \beta_n)$$

respecto de esta base, tendremos, de acuerdo con (28.3),

$$|x| = (\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2)^{1/2}.$$

Por consiguiente,

$$\cos(x, y) = \frac{\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n}{(\alpha_1^2 + \ldots + \alpha_n^2)^{1/2}(\beta_1^2 + \ldots + \beta_n^2)^{1/2}}.$$

La analogía total con las fórmulas (25.4), (25.5) es evidente.

De este modo, las nociones introducidas de longitud, ángulo y distancia concuerdan enteramente con las nociones análogas en el espacio de segmentos dirigidos.

Ejercicios.

1. Demuéstrese que la longitud de una suma de cualquier número de vectores no es superior a la suma de longitudes de estos vectores.
2. Demuéstrese que el cuadrado de la longitud de una suma de cualquier número de vectores ortogonales es igual a la suma de cuadrados de las longitudes de dichos vectores.
3. En un espacio euclídeo de polinomios, que dependen de una variable t, hállense los ángulos del triángulo formado por los vectores $1,\ t^2,\ 1 - t^3$.
4. ¿Cuál es la distancia entre los polinomios $3t^2 + 6$ y $2t^2 + t + 1$?
5. Demuéstrese que un triángulo en el espacio euclídeo es rectángulo cuando, y sólo cuando, la longitud de un lado es igual al producto de la longitud del otro lado por el coseno del ángulo entre ellos.

§ 30. Línea oblicua, perpendicular, proyección

Antes de extender las nociones de oblicua, perpendicular y proyección a los espacios euclídeos abstractos, considerémoslas en un espacio de segmentos dirigidos.



Fig. 30.1

Sea dado un plano L. Desde un punto M tracemos una perpendicular al plano y designemos con M_L su base (fig. 30.1). Con el fin de comunicar al problema una interpretación vectorial, elijamos en el plano L un punto O y consideremos el espacio V_3 de segmentos dirigidos sujetados al punto O. El plano L forma un subespacio, por lo cual la construcción de la perpendicular bajada desde el punto M al plano L se reduce a la descomposición del vector $\overrightarrow{OM}$ del espacio en la suma

$$\overrightarrow{OM} = \overrightarrow{OM_L} + \overrightarrow{M_LM}, \tag{30.1}$$

donde $\overrightarrow{OM_L} \in L$ y $\overrightarrow{M_LM} \perp L$. De los razonamientos geométricos se ve que la descomposición (30.1) existe siempre y es única.

El ejemplo examinado sugiere cómo, en el caso general, debe plantearse el problema sobre una perpendicular. Supongamos que en el espacio euclídeo E se ha fijado cierto subespacio L. Tomemos un vector arbitrario f de E e investiguemos la posibilidad de su descomposición en la suma

$$f = g + h, \tag{30.2}$$

donde $g \in L$ y $h \perp L$.

Con este problema ya nos encontramos anteriormente. Efectivamente, la condición $h \perp L$ es equivalente a la condición $h \in L^{\perp}$,

De acuerdo con el teorema 28.2, el espacio euclídeo E es la suma directa de los subespacios L y $L^{\perp}$. Por esta razón, la descomposición (30.2) siempre existe y es única.

Al tomar en consideración la analogía con la descomposición (30.1), el vector g en la descomposición (30.2) se llamará *proyección del vector f sobre el subespacio L*; h, *perpendicular trazada desde el vector f al subespacio L*; el propio vector f, *línea oblicua al subespacio L*.

Se sabe que en la geometría elemental la longitud de una perpendicular nunca supera la longitud de una línea oblicua. Una situación análoga tiene lugar también en el espacio euclídeo. Los vectores g, h en la descomposición (30.2) son ortogonales. Por ello, de acuerdo con el teorema de Pitágoras,

$$|f|^2 = |g|^2 + |h|^2,$$

de donde se deduce que

$$|h| \leqslant |f|.$$

Está claro que la longitud de la perpendicular h al subespacio L equivale a la longitud de la línea oblicua f al mismo subespacio cuando, y sólo cuando, $f \perp L$.

Al problema sobre la perpendicular se le puede dar también otra interpretación. Consideraremos otra vez un vector arbitrario f de E. Este vector no pertenece forzosamente al subespacio L. Por consiguiente, se puede plantear el problema sobre la obtención en L de un vector cuya disposición fuera más próxima a f en el sentido de la distancia introducida anteriormente.

Tomemos un vector arbitrario z de L. Al sustraerlo de ambos miembros de la igualdad (30.2) obtendremos

$$f - z = (g - z) + h.$$

Ya que el vector h es ortogonal al vector $g - z$, de acuerdo con el teorema de Pitágoras, tenemos

$$|f - z|^2 = |g - z|^2 + |h|^2.$$

Por esto

$$|f - z| \geqslant h,$$

con la particularidad de que la igualdad es verdadera cuando, y sólo cuando, $z = g$.

Así pues, de todos los vectores del subespacio L la proyección del vector f sobre L es la más próxima al vector f. Esto es testimonio de que

$$\rho(f, L) = \rho(f, g).$$

Por analogía con los segmentos dirigidos llamemos *ángulo entre el vector f y el subespacio L* al menor de los ángulos formados por el vector f y los vectores z de L. Tomando en consideración la desigual-

dad de Cauchy—Buniakovski y la descomposición (30.2), hallemos

$$\cos(f, z) = \frac{(f, z)}{|f|\,|z|} = \frac{(z+h, z)}{|f|\,|z|} = \frac{(z, z)}{|f|\,|z|} \leqslant \frac{|z|}{|f|}.$$

Es evidente que esta desigualdad se convierte en una igualdad cuando, y sólo cuando, los vectores z y g forman un ángulo nulo.

De este modo, el ángulo entre el vector f y el subespacio L coincide con el ángulo formado por el vector f y la proyección de éste sobre el subespacio L.

Las propiedades citadas que poseen la perpendicular y la proyección reflejan el aspecto geométrico de estas nociones. Ahora considerémoslas desde el punto de vista algebraico. Con el subespacio L fijado, todo vector f del espacio euclídeo E define unívocamente respecto de L sus dos componentes. Por consiguiente, se puede considerar que la descomposición (30.2) prefija dos funciones

$$g = \mathrm{pr}_L f,$$
$$h = \mathrm{ort}_L f.$$

Como argumentos de la función puede figurar cualquier vector de E; como valores de la función $\mathrm{pr}_L f$ sirve un vector de L; como valores de la función $\mathrm{ort}_L f$, un vector de $L^{\perp}$.

En vista de la correlación $(L^{\perp})^{\perp} = L$, la perpendicular y la proyección están ligadas mediante las igualdades

$$\mathrm{pr}_L f = \mathrm{ort}_{L^{\perp}} f,$$
$$\mathrm{ort}_L f = \mathrm{pr}_{L^{\perp}} f. \tag{30.3}$$

Por ello, el estudio de estas funciones siempre se reduce, de hecho, al estudio de una de ellas.

Tomemos dos vectores arbitrarios x, y de E. Conforme a la descomposición (30.2), se tiene

$$x = \mathrm{pr}_L x + \mathrm{ort}_L x, \tag{30.4}$$
$$y = \mathrm{pr}_L y + \mathrm{ort}_L y.$$

Al sumar estas igualdades término a término y al multiplicar la primera de ellas por un número real arbitrario λ, obtenemos

$$x + y = (\mathrm{pr}_L x + \mathrm{pr}_L y) + (\mathrm{ort}_L x + \mathrm{ort}_L y),$$
$$\lambda x = (\lambda\, \mathrm{pr}_L x) + (\lambda\, \mathrm{ort}_L x).$$

Por comprobación inmediata nos convencemos de que los vectores dentro del primer paréntesis pertenecen a L, mientras que los encerrados dentro del segundo paréntesis son perpendiculares a L. Debido a la unicidad de las descomposiciones del tipo (30.2) esto significa

la validez de las siguientes correlaciones

$$\begin{aligned} \operatorname{pr}_L (x+y) &= \operatorname{pr}_L x + \operatorname{pr}_L y, \\ \operatorname{pr}_L \lambda x &= \lambda \operatorname{pr}_L x \end{aligned} \qquad (30.5)$$

pero la función pr_L y, por supuesto, de las correlaciones análogas

$$\begin{aligned} \operatorname{ort}_L (x+y) &= \operatorname{ort}_L x + \operatorname{ort}_L y, \\ \operatorname{ort}_L (\lambda x) &= \lambda \operatorname{ort}_L x \end{aligned} \qquad (30.6)$$

para la función ort_L. Aquí tiene lugar una completa coincidencia de las fórmulas (25.9) y (30.5).

Observemos que $\operatorname{ort}_L z = 0$ para todo vector z de L. Por eso, de la primera igualdad (30.6) se deduce que

$$\operatorname{ort}_L (x+z) = \operatorname{ort}_L (x).$$

Por consiguiente, el valor de la función ort_L no se altera, si al argumento se adiciona cualquier vector del subespacio L. En particular, si tomamos $z = -\operatorname{pr}_L x$, entonces, teniendo en cuenta (30.4), obtenemos

$$\operatorname{ort}_L (\operatorname{ort}_L x) = \operatorname{ort}_L x. \qquad (30.7)$$

Una correlación análoga tiene lugar también para la proyección. A saber,

$$\operatorname{pr}_L (\operatorname{pr}_L x) = \operatorname{pr}_L x. \qquad (30.8)$$

Supongamos que el subespacio L es la suma ortogonal de los subespacios L_1 y L_2. Elijamos un vector arbitrario x de E y representémoslo en forma de la suma

$$x = (\operatorname{pr}_{L_1} x + \operatorname{pr}_{L_2} x) + (x - \operatorname{pr}_{L_1} x - \operatorname{pr}_{L_2} x).$$

El vector dentro del primer paréntesis pertenece, evidentemente, al subespacio $L_1 \oplus L_2$. El vector dentro del segundo paréntesis es ortogonal a $L_1 \oplus L_2$, de lo que es fácil convencerse transformándolo con ayuda de las correlaciones (30.4) del modo siguiente

$$x - \operatorname{pr}_{L_1} x - \operatorname{pr}_{L_2} x = \operatorname{ort}_{L_1} x - \operatorname{pr}_{L_2} x = \operatorname{ort}_{L_2} x - \operatorname{pr}_{L_1} x. \qquad (30.9)$$

Concluimos, a partir de lo dicho, que

$$\operatorname{pr}_{L_1 \oplus L_2} x = \operatorname{pr}_{L_1} x + \operatorname{pr}_{L_2} x.$$

Una perpendicular bajada desde el vector x sobre el subespacio $L_1 \oplus L_2$ es igual a una de las expresiones (30.9). Si, en particular, $x \perp L_1$, entonces

$$\operatorname{ort}_{L_1 \oplus L_2} x = \operatorname{ort}_{L_2} x. \qquad (30.10)$$

Ejercicios.

1. ¿Tendrá lugar en un espacio euclídeo el análogo del teorema sobre las tres perpendiculares?

2. Demuéstrese que la suma de dos ángulos, formados por el vector f y los subespacios L y $L^{\perp}$, es igual a $\pi/2$.

3. Hállense la perpendicular y la proyección del vector f sobre los subespacios triviales.

4. Demuéstrese que si para los espacios fijados L_1, L_2 y para cualquier vector x se verifica la igualdad

$$\mathrm{pr}_{L_1+L_2}\, x = \mathrm{pr}_{L_1}\, x + \mathrm{pr}_{L_2}\, x,$$

la suma $L_1 + L_2$ será ortogonal.

5. Demuéstrese que si los subespacios $L_1, L_2, \ldots, L_m$ son ortogonales dos a dos, entonces para todo vector x de E

$$|x|^2 \geqslant \sum_{i=1}^{m} |\mathrm{pr}_{L_i}\, x|^2.$$

§ 31. Isomorfismo euclídeo

En el transcurso de nuestras investigaciones ya hemos observado, más de una vez, que las propiedades de un espacio euclídeo abstracto coinciden con las de los espacios de segmentos dirigidos. Se podría continuar extendiendo los hechos y teoremas de la geometría elemental al espacio euclídeo. Sin embargo, no hay necesidad en esto.

Introduzcamos la noción de isomorfismo euclídeo. Diremos que los espacios euclídeos E y E' son *isomorfos según Euclides*, si son isomorfos como espacios lineales reales y, además, para todo par de vectores x, y de E y los vectores correspondientes x', y' de E' se verifica la igualdad

$$(x, y) = (x', y').$$

TEOREMA 31.1. *Para que dos espacios euclídeos sean isomorfos según Euclides, es necesario y suficiente que sean iguales sus dimensiones.*

DEMOSTRACIÓN. Si dos espacios euclídeos E y E' son isomorfos según Euclides, serán isomorfos también como espacios reales lineales. Pero, los espacios lineales de tal índole son de igual dimensión.

Consideremos ahora dos espacios euclídeos E y E' de una misma dimensión n. Sea $e_1, e_2, \ldots, e_n$ la base ortonormalizada en E y sea $e_1', e_2', \ldots, e_n'$ la base ortonormalizada en E'. A todo vector

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n$$

del espacio E ponemos en correspondencia el vector

$$x' = \alpha_1 e_1' + \alpha_2 e_2' + \ldots + \alpha_n e_n'$$

del espacio E'. De acuerdo con lo demostrado anteriormente, esta correspondencia representa un isomorfismo. Tomemos ahora otro

par de vectores correspondientes de E y E'

$$y = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n,$$
$$y' = \beta_1 e_1' + \beta_2 e_2' + \ldots + \beta_n e_n'.$$

Según (28.3) tenemos

$$(x, y) = \alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n = (x', y').$$

El teorema queda demostrado.

Nos interesan siempre sólo aquellas propiedades de los espacios lineales que son consecuencias de las operaciones fundamentales que actúan en los espacios. Desde este punto de vista los espacios isomorfos según Euclides tienen propiedades iguales. Por ello, cualquier teorema geométrico, demostrado para el espacio V_3, será verídico también en cualquier subespacio tridimensional del espacio euclídeo. Por consiguiente, será válido también en todo espacio euclídeo. Por supuesto, como espacio euclídeo tipo puede servir el espacio aritmético R_n con un producto escalar introducido conforme a (27.2).

Ejercicios.

1. Constrúyase un isomorfismo euclídeo entre los espacios V_3 y R_3.
2. Demuéstrese que en los espacios isomorfos según Euclides un sistema ortonormalizado de vectores pasa a ser de nuevo un sistema ortonormalizado.
3. Demuéstrese que en los espacios isomorfos según Euclides los ángulos entre los pares de correspondientes vectores son iguales.
4. Demuéstrese que en los espacios isomorfos según Euclides una perpendicular y una proyección pasan a ser una perpendicular y una proyección, respectivamente.

§ 32. Espacio unitario

Los conceptos métricos fundamentales han sido extendidos sólo a los espacios lineales reales. Resultados análogos tienen lugar también en un espacio lineal complejo.

El espacio lineal complejo U se denomina *unitario*, si a todo par de vectores x, y de U se le ha puesto en correspondencia un número complejo (x, y), llamado *producto escalar*, con la particularidad de que se cumplen los siguientes axiomas:

1) $(x, y) = \overline{(y, x)}$,
2) $(\lambda x, y) = \lambda (x, y)$,
3) $(x + y, z) = (x, z) + (y, z)$,
4) $(x, x) > 0$ cuando $x \neq 0$; $(0, 0) = 0$

para unos vectores arbitrarios x, y, z de U y un número complejo arbitrario λ.

La raya en el primer axioma es un signo de conjugación compleja. Esta única diferencia de los axiomas del espacio euclídeo no lleva

consigo distinciones profundas, sin embargo no se debe olvidarlo. Así por ejemplo, si en un espacio euclídeo tiene lugar la igualdad $(x, \lambda y) = \lambda (x, y)$, en el unitario se verificará la igualdad $(x, \lambda y) = \bar{\lambda} (x, y)$.

En el espacio unitario U se pueden introducir algunos conceptos métricos. Al igual que en el caso real, llamemos longitud del vector la magnitud

$$|x| = +(x, x)^{1/2}.$$

Todo vector no nulo tiene una longitud positiva, la longitud de un vector nulo es igual a cero. Para cualquier λ complejo se verifica la correlación

$$|\lambda x| = |\lambda| \cdot |x|.$$

Es válida también la desigualdad de Cauchy—Buniakovski

$$|(x, y)|^2 \leq (x, x)(y, y).$$

La demostración se lleva a cabo siguiendo el mismo esquema que en el caso real.

En el espacio unitario el concepto de ángulo entre los vectores, por regla general, no se introduce. Se considera solamente el caso en que los vectores x e y son ortogonales. Al igual que en el caso real, se entiende en tal circunstancia el cumplimiento de la igualdad

$$(x, y) = 0.$$

Es evidente que $(y, x) = \overline{(x, y)} = 0$.

En esencia, toda la teoría del espacio euclídeo, examinada más arriba, se aplica a un espacio unitario sin que cambien las definiciones y los esquemas generales de las demostraciones.

A título del espacio unitario puede servir el espacio aritmético C_n, siempre que para los vectores

$$x = (\alpha_1, \alpha_2, \ldots, \alpha_n),$$
$$y = (\beta_1, \beta_2, \ldots, \beta_n)$$

el producto escalar se introduzca de la manera siguiente:

$$(x, y) = \sum_{i=1}^{n} \alpha_i \bar{\beta}_i. \qquad (32.1)$$

Con el ejemplo de este espacio se muestra fácilmente la significación de la conjugación compleja en el primer axioma. Si en el espacio C_n introdujéramos el producto escalar según la fórmula (27.2), entonces, en el espacio C_3, por ejemplo, para el vector

$$x = (3, 4, 5i)$$

tendríamos

$$(x, x) = 9 + 16 + 25i^2 = 0.$$

El cuarto axioma, que es muy importante, resultaría incumplido.

Ejercicios.

1. Compárense el espacio euclídeo R_n y el espacio unitario C_n.
2. Escríbase la desigualdad de Cauchy—Buniakovski en el espacio C_n.
3. Si en un espacio complejo el producto escalar se introduce de acuerdo con los axiomas (27.1), ¿podrá cumplirse en tal espacio la desigualdad de Cauchy—Buniakovski?
4. Si en un espacio complejo el producto escalar se introduce de acuerdo con los axiomas (27.1), ¿podrá existir en tal espacio una base ortogonal?

§ 33. Dependencia lineal y sistemas ortonormalizados

Ya hemos visto en el § 22 que la independencia lineal de un sistema de vectores de una base puede perturbarse como resultado de *pequeñas* alteraciones en los mismos vectores. Este fenómeno conduce a complicaciones bastante grandes cuando el concepto de base se emplea en la resolución de los problemas prácticos. No obstante, conviene subrayar que no todas las bases poseen esta desfavorable propiedad. En particular, *no la tiene* ninguna base ortonormalizada.

Supongamos que en un espacio euclídeo o en un espacio unitario se ha elegido una base ortonormalizada arbitraria $e_1, e_2, \ldots, e_n$. Si para cierto vector b tiene lugar la descomposición

$$b = \sum_{i=1}^{n} \alpha_i e_i,$$

entonces, según (28.4)

$$|b|^2 = \sum_{i=1}^{n} |\alpha_i|^2. \qquad (33.1)$$

Examinaremos ahora el sistema de vectores $e_1 + \varepsilon_1, e_2 + \varepsilon_2, \ldots, e_n + \varepsilon_n$ y supondremos que el sistema es linealmente dependiente. Esto significa que existen tales números $\beta_1, \beta_2, \ldots, \beta_n$, no nulos a la vez, que

$$\sum_{i=1}^{n} \beta_i (e_i + \varepsilon_i) = 0.$$

De aquí se desprende que

$$\sum_{i=1}^{n} \beta_i e_i = -\sum_{i=1}^{n} \beta_i \varepsilon_i.$$

Al hacer uso de la igualdad (33.1) y desigualdad (27.6), tenemos

$$\sum_{i=1}^{n} |\beta_i|^2 = \Big|\sum_{i=1}^{n} \beta_i e_i\Big|^2 = \Big|\sum_{i=1}^{n} \beta_i \varepsilon_i\Big|^2 \leqslant$$

$$\leqslant \Big(\sum_{i=1}^{n} |\beta_i|\,|\varepsilon_i|\Big)^2 \leqslant \Big(\sum_{i=1}^{n} |\beta_i|^2\Big)\Big(\sum_{i=1}^{n} |\varepsilon_i|^2\Big).$$

Comparando los miembros primero y segundo de las correlaciones obtenidas, llegamos a la conclusión que

$$\sum_{i=1}^{n} |\varepsilon_i|^2 \geqslant 1.$$

De este modo, la desigualdad obtenida significa que al cumplirse la condición

$$\sum_{i=1}^{n} |\varepsilon_i|^2 < 1, \qquad (33.2)$$

el sistema de vectores

$$e_1 + \varepsilon_1,\ e_2 + \varepsilon_2,\ \ldots,\ e_n + \varepsilon_n$$

será, a ciencia cierta, linealmente independiente.

La peculiaridad de los sistemas ortonormalizados que acabamos de indicar ha determinado el amplio uso de estos sistemas en la construcción de los más diversos algoritmos de cómputo relacionados con la descomposición según la base.

Ejercicios.

1. Sea $e_1, e_2, \ldots, e_n$ una base ortogonal de un espacio euclídeo. Demuéstrese que el sistema de vectores $x_1, x_2, \ldots, x_n$ es linealmente independiente, si

$$\sum_{i=1}^{n} \cos(x_i, e_i) > n - \frac{1}{2}.$$

2. Supongamos que los vectores $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ para $i = 1, 2, \ldots, n$ están definidos por medio de sus coordenadas en una base arbitraria. Demuéstrese que si

$$|x_{ii}|^2 > n \sum_{k \neq i} |x_{ik}|^2$$

para cualquier valor de i, el sistema $x_1, x_2, \ldots, x_n$ es linealmente independiente.

CAPITULO 4 VOLUMEN DEL SISTEMA DE VECTORES EN UN ESPACIO LINEAL

§ 34. Productos vectorial y mixto

De nuevo comencemos nuestras investigaciones por un espacio de segmentos dirigidos. Supongamos, como siempre, que se ha fijado cierto sistema rectangular cartesiano de coordenadas cuyo origen se ubica en O y la base es i, j, k.

Tres vectores se denominan *terna*, si se ha indicado cuál de estos vectores es el primero, cuál es el segundo y cuál el tercero. Al escribir una terna de vectores, dispondremos los propios vectores de izquierda a derecha en el orden de seguimiento.

Una terna de vectores no coplanares a, b, c se llama *derecha* (*izquierda*), si dichos vectores se disponen igual que los dedos pulgar, índice no doblado y del corazón, respectivamente, de la mano *derecha* (*izquierda*).

De cualesquiera tres vectores no coplanares a, b, c se pueden componer las siguientes seis ternas

$$abc,\ bca,\ cab,\ bac,\ acb,\ cba.$$

Las primeras tres ternas son de la misma denominación que la terna abc, las demás ternas son de denominación contraria. Cabe notar que si en una terna se cambian de lugares cualesquiera dos vectores, dicha terna cambiará su denominación.

Un sistema de coordenadas, afín o cartesiano, se llama *derecho* (*izquierdo*), si los vectores básicos forman una terna derecha (izquierda). Hasta el presente momento nuestras investigaciones no dependían de la denominación que tenía la base del sistema de coordenadas. Ahora en las investigaciones aparecerán ciertas diferencias. Es por eso que para concretar consideraremos en lo sucesivo solamente sistemas de coordenadas derechos.

Sean dados dos vectores no colineales a, b. Pondremos en correspondencia a estos vectores el tercer vector c que satisfaga las siguientes condiciones:

1) el vector c es ortogonal a cada uno de los vectores a, b,
2) la terna abc es derecha,

3) la longitud del vector c es igual numéricamente al área S del paralelogramo construido en los vectores a, b, reducidos a un origen común. Si los vectores a, b son colineales, a tal par de vectores le pondremos en correspondencia el vector nulo.

La correspondencia construida es una operación algebraica en el espacio V_3. Se llama *multiplicación vectorial de los vectores* a, b y se indica con el símbolo

$$c = [a, b].$$

Consideremos los vectores básicos i, j, k. Por definición del producto vectorial tendremos

$$\begin{aligned} &[i, i] = 0, && [i, j] = k, && [i, k] = -j, \\ &[j, i] = -k, && [j, j] = 0, && [j, k] = i, \\ &[k, i] = j, && [k, j] = -i, && [k, k] = 0. \end{aligned} \qquad (34.1)$$

De estas correlaciones se deduce, en particular, que la operación del producto vectorial no es conmutativa.

Toda terna abc de vectores no coplanares aplicados a un punto común O define cierto paralelepípedo. El punto O es uno de los vértices y los vectores a, b, c son las aristas. Designaremos el volumen de este paralelepípedo mediante el símbolo $V(a, b, c)$, subrayando de esta manera, que el volumen depende de los vectores a, b, c. Si la terna a, b, c es coplanar, el volumen se considerará igual a cero. Atribuyamos al volumen el signo más, si la terna no coplanar abc es derecha y el signo menos, si es izquierda. Este nuevo concepto, definido de tal modo, se denominará *volumen orientado* del paralelepípedo y se indicará con el símbolo $V_{\pm}(a, b, c)$.

El volumen y el volumen orientado pueden considerarse como ciertas funciones numéricas de tres argumentos vectoriales que toman determinados valores reales para cada terna de vectores abc. El volumen es siempre no negativo, mientras que el volumen orientado puede tener un signo cualquiera. En lo que sigue se pondrá de manifiesto el sentido bien determinado que radica en la distinción de estas nociones.

Sean dados tres vectores arbitrarios a, b, c. Si a se multiplica vectorialmente a la derecha por b y luego el vector $[a, b]$ se multiplica escalarmente por c, el número que se obtiene $([a, b], c)$ se llama *producto mixto de los vectores* a, b, c.

TEOREMA 34.1. *Un producto mixto* $([a, b], c)$ *es igual al volumen orientado del paralelepípedo construido sobre los vectores* a, b, c *reducidos a un origen común.*

DEMOSTRACIÓN. Sin restringir la generalidad podemos considerar que los vectores a, b no son colineales, puesto que en el caso contrario $[a, b] = 0$ y la afirmación del teorema resulta evidente. Sea, como antes, S el área del paralelogramo construido sobre los vectores a, b.

Conforme a (26.4) se tiene

$$([a, b], c) = |[a, b]|\,\{\mathrm{pr}_{[a,b]}\, c\} = S\,\{\mathrm{pr}_{[a,b]}\, c\}. \qquad (34.2)$$

Supongamos que los vectores a, b, c son no coplanares. Entonces, $\{\mathrm{pr}_{[a,b]}\, c\}$ es igual, salvo el signo, a la altura h del paralelepípedo construido sobre los vectores a, b, c trasladados al origen común, a condición de que de base sirve el paralelogramo construido sobre los vectores a, b (fig. 34.1). De este modo, el segundo miembro de (34.2) es igual, salvo el signo, al volumen del paralelepípedo construido sobre los vectores a, b, c.

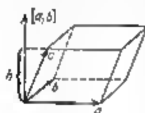


Fig. 34.1.

Evidentemente, $\{\mathrm{pr}_{[a,b]}\, c\} = +h$, si los vectores $[a, b]$ y c se disponen por un lado respecto al plano definido por los vectores a, b. Mas, en este caso, la terna abc también es derecha. En el caso contrario $\{\mathrm{pr}_{[a,b]}\, c\} = -h$. Si los vectores abc son coplanares, entonces c se dispone en el plano definido por los vectores a, b por lo cual $\{\mathrm{pr}_{[a,b]}\, c\} = 0$. El teorema queda demostrado.

COROLARIO. *Para cualesquiera tres vectores a, b, c se verifica la correlación*

$$([a, b], c) = (a, [b, c]). \qquad (34.3)$$

Efectivamente, de la propiedad de simetría del producto escalar se desprende que $(a, [b, c]) = ([b, c], a)$, por lo cual basta señalar que $([a, b], c) = ([b, c], a)$. Mas, la última igualdad es evidente, puesto que las ternas abc y bca son de una misma denominación y a ellas corresponde un mismo paralelepípedo.

La correlación (34.3) permite realizar con gran eficacia las investigaciones algebraicas. Demostremos primeramente que para cualesquiera vectores a, b, c y para todo número real α tienen lugar las siguientes propiedades de la multiplicación vectorial:

1) $[a, b] = -[b, a]$,
2) $[\alpha a, b] = \alpha [a, b]$,
3) $[a + b, c] = [a, c] + [b, c]$,
4) $[a, a] = 0$.

La propiedad 4 proviene de forma evidente de la definición. Con el fin de demostrar las propiedades restantes, haremos uso del hecho que los vectores x e y son iguales entre sí cuando, y sólo cuando,

$$(x, d) = (y, d)$$

para todo vector d.

Sea d un vector arbitrario. Las ternas abd y bad son de diferentes denominaciones. Por consiguiente, en virtud del teorema 34.1 y las propiedades del producto escalar, concluimos que

$$([a, b], d) = -([b, a], d) = (-[b, a], d).$$

Como d es un vector arbitrario, esto es un indicio de que $[a, b] = -[b, a]$ y la primera propiedad queda demostrada.

Para demostrar las propiedades segunda y tercera procedemos de manera análoga, tomando en consideración, además, la correlación (34.3). Tenemos

$$([\alpha a, b], d) = (\alpha a, [b, d]) = \alpha (a, [b, d]) =$$
$$= \alpha ([ab], d) = (\alpha [a, b], d),$$

lo que significa la validez de la propiedad 2. Luego,

$$([a + b, c], d) = (a + b, [c, d]) = (a, [c, d]) + (b, [c, d]) =$$
$$= ([a, c], d) + ([b, c], d) = ([a, c] + [b, c], d)$$

y la propiedad 3 resulta también lícita. Respecto del segundo factor tienen lugar las igualdades correspondientes:

$$[a, \alpha b] = -[\alpha b, a] = -\alpha [b, a] = \alpha [a, b],$$
$$[a, b + c] = -[b + c, a] = -[b, a] - [c, a] = [a, b] + [a, c].$$

Ahora podemos investigar las propiedades algebraicas de un volumen orientado considerándolo como una función definida en las ternas de vectores. Sea, por ejemplo, el vector a una combinación lineal de ciertos vectores a', a''. En este caso

$$([\alpha a' + \beta a'', b], c) = (\alpha [a', b] + \beta [a'', b], c) =$$
$$= \alpha ([a', b], c) + \beta ([a'', b], c).$$

Por consiguiente,

$$V^{\pm} (\alpha a' + \beta a'', b, c) = \alpha V^{\pm} (a', b, c) + \beta V^{\pm} (a'', b, c)$$

para cualesquiera vectores a', a'' y números reales α, β.

Cuando dos argumentos se cambian de lugar, el volumen orientado sólo cambia de signo, a consecuencia de lo cual la propiedad análoga referente a la combinación lineal es verídica para cada argumento. Teniendo presente precisamente esta propiedad, diremos que el volumen orientado representa en sí *una función lineal* respecto de todo argumento.

Si los vectores abc son linealmente dependientes, serán coplanares y, por ende, el volumen orientado en el caso dado es igual a cero. Luego, al tomar en consideración las correlaciones (34.1) hallamos que

$$V^{\pm} (i, j, k) = ([i, j], k) = (k, k) = 1.$$

Así pues, podemos concluir que el volumen orientado posee, en su calidad de función, las siguientes propiedades:

A) el volumen orientado es una función lineal respecto de todo argumento,

B) el volumen orientado es igual a cero en todos los sistemas linealmente dependientes, (34.4)

C) el volumen orientado es igual a la unidad al menos en un sistema ortonormalizado fijado de vectores.

Por supuesto, estamos lejos de haber enunciado todas las propiedades del volumen orientado. Las propiedades destacadas en (34.4), como también las otras, pueden establecerse con facilidad, si se conoce la expresión explícita de los productos vectorial y mixto en términos de las coordenadas de los vectores a, b, c.

TEOREMA 34.2. *Si los vectores a, b están dados mediante sus coordenadas rectangulares cartesianas*

$$a = (x_1,\ y_1,\ z_1),$$
$$b = (x_2,\ y_2,\ z_2),$$

el producto vectorial tendrá las coordenadas siguientes:

$$[a,\ b] = (y_1 z_2 - y_2 z_1,\ z_1 x_2 - z_2 x_1,\ x_1 y_2 - x_2 y_1). \qquad (34.5)$$

DEMOSTRACIÓN. Teniendo en cuenta que las coordenadas de los vectores prefijados determinan las descomposiciones

$$a = x_1 i + y_1 j + z_1 k,$$
$$b = x_2 i + y_2 j + z_2 k,$$

y apoyándonos en las propiedades algebraicas del producto vectorial, hallamos

$$\begin{aligned}[a,\ b] = x_1 x_2 [i,\ i] &+ x_1 y_2 [i,\ j] + x_1 z_2 [i,\ k] + \\ &+ y_1 x_2 [j,\ i] + y_1 y_2 [j,\ j] + y_1 z_2 [j,\ k] + \\ &+ z_1 x_2 [k,\ i] + z_1 y_2 [k,\ j] + z_1 z_2 [k,\ k].\end{aligned}$$

La validez de la afirmación del teorema se desprende ahora de la correlación (34.1).

COROLARIO. *Si el vector c está dado también mediante las coordenadas x_3, y_3, z_3 en el mismo sistema cartesiano, entonces*

$$\begin{aligned}([a,\ b],\ c) = x_1 y_2 z_3 &+ x_2 y_3 z_1 + x_3 y_1 z_2 - \\ &- x_3 y_2 z_1 - x_2 y_1 z_3 - x_1 y_3 z_2. \qquad (34.6)\end{aligned}$$

La introducción del volumen orientado y la investigación de sus propiedades algebraicas permiten llegar a deducciones de importancia referentes a la longitud, el área y el volumen.

Observamos que las bases derechas e izquierdas determinan la partición del conjunto de todas las bases del espacio en dos clases. La propia denominación «derechas e izquierdas» no lleva ningún sentido profundo, sino que está relacionada con un método cómodo para distinguir la clase a que pertenece tal o cual base. A estas dos clases está ligado, en esencia, el propio concepto de volumen orientado.

Con hechos de tal índole ya nos hemos encontrado. Todas las bases en una recta se pueden dividir también en dos clases, reuniendo en una clase los vectores dirigidos hacia un mismo lado. En este caso resulta que la magnitud de un segmento dirigido es un análogo completo del volumen orientado, siempre que ambos conceptos se consideren como funciones de los sistemas de vectores. La propiedad **A** tiene lugar de acuerdo con las correlaciones (9.8). La propiedad **B** es válida, puesto que la magnitud del segmento nulo es igual a cero. El cumplimiento de la propiedad **C** es obvio.

Una investigación análoga independiente se podría realizar también en el caso de un plano. Sin embargo, resulta más ventajoso hacer uso de los resultados ya obtenidos anteriormente. Fijemos un sistema rectangular cartesiano de coordenadas Oxy. Completémoslo hasta obtener el sistema derecho de coordenadas $Oxyz$ en el espacio. Prestemos atención a que según sea la disposición de los ejes Ox y Oy, el eje Oz pueda tener una de las dos direcciones posibles. Esto también determina la partición del conjunto de las bases de un plano en dos clases. El área orientada $S^{\pm}(a, b)$ del paralelogramo, construido sobre los vectores a, b en el plano Oxy, puede ser definida, por ejemplo, mediante la igualdad $S^{\pm}(a, b) = V^{\pm}(a, b, k)$. Naturalmente, las propiedades **A**, **B**, **C** subsisten en este caso también.

De este modo, atribuyendo a las longitudes, las áreas y los volúmenes ciertos signos y considerando dichos conceptos como funciones definidas en los sistemas de vectores, podemos conseguir que todas estas funciones tengan las mismas propiedades algebraicas **A**, **B**, **C** de (34.4).

Ejercicios.

1. Demuéstrese que los vectores a, b, c son coplanares cuando, y sólo cuando, su producto mixto es igual a cero.
2. Demuéstrese que para cualesquiera tres vectores a, b, c se verifica la correlación

$$[a, [b, c]] = (a, c)\, b - (a, b)\, c.$$

3. Demuéstrese que la multiplicación vectorial no es una operación asociativa.
4. Hállese la expresión del área orientada de un paralelogramo en términos de las coordenadas cartesianas de vectores en un plano.
5. ¿Cambiarán las fórmulas (34.5), (34.6), si el sistema de coordenadas, respecto del cual están dados los vectores, es izquierdo?

§ 35. Volumen y volumen orientado del sistema de vectores

En los espacios lineales de segmentos dirigidos el área y el volumen son las nociones que se derivan de la longitud del segmento. El concepto de longitud ya lo hemos extendido al espacio euclídeo abstracto. Ahora examinemos un problema análogo en relación al área y el volumen.

Supongamos que en un plano están dados dos vectores no colineales x_1, x_2. Construyamos sobre estos vectores un paralelogramo, tomando por base el vector x_1 (fig. 35.1). Empleando el extremo del vector x_2 como punto de partida, tracemos la perpendicular h a la base. El área $S(x_1, x_2)$ del paralelogramo se expresará mediante la fórmula

$$S(x_1, x_2) = |x_1|\,|h|. \qquad (35.1)$$

Designaremos con L_0 el subespacio nulo y con L_1, la cápsula lineal construida sobre el vector x_1. Puesto que

$$|x_1| = |\mathrm{ort}_{L_0} x_1|,$$

la fórmula (35.1) puede ser escrita en la forma siguiente:

$$S(x_1, x_2) = |\mathrm{ort}_{L_0} x_1|\,|\mathrm{ort}_{L_1} x_2|. \qquad (35.2)$$

Luego, tomemos en el espacio tres vectores no coplanares x_1, x_2, x_3. Construyamos un paralelepípedo sobre estos vectores, tomando por



Fig. 35.1.

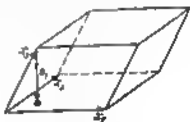


Fig. 35.2.

base el paralelogramo formado por los vectores x_1, x_2 (fig. 35.2). Desde el extremo del vector x_3 tracemos sobre la base la perpendicular h_1. El volumen $V(x_1, x_2, x_3)$ del paralelepípedo se expresará mediante la fórmula

$$V(x_1, x_2, x_3) = S(x_1, x_2)\,|h_1|.$$

Al designar con L_2 la cápsula lineal construida sobre los vectores x_1, x_2, tendremos, conforme a (35.2),

$$V(x_1, x_2, x_3) = |\mathrm{ort}_{L_0} x_1|\,|\mathrm{ort}_{L_1} x_2|\,|\mathrm{ort}_{L_2} x_3|.$$

De este modo, la longitud del vector, el área del paralelogramo y el volumen del paralelepípedo en los espacios lineales V_1, V_2, V_3 se expresan mediante las fórmulas en las cuales no es difícil descubrir cierta regularidad:

$$\begin{gathered} |x_1| = |\mathrm{ort}_{L_0} x_1|, \\ S(x_1, x_2) = |\mathrm{ort}_{L_0} x_1|\,|\mathrm{ort}_{L_1} x_2|, \\ V(x_1, x_2, x_3) = |\mathrm{ort}_{L_0} x_1|\,|\mathrm{ort}_{L_1} x_2|\,|\mathrm{ort}_{L_2} x_3|. \end{gathered} \qquad (35.3)$$

En particular, en todo caso el número de vectores coincide con la dimensión del espacio.

Estas fórmulas dicen cómo se debe introducir el concepto de volumen en un espacio euclídeo E_n de dimensión n. Supongamos que en E_n se ha dado un sistema arbitrario de vectores $x_1, x_2, \ldots, x_n$. Designemos por L_0 el subespacio nulo y por L_i, la cápsula lineal formada por los vectores $x_1, \ldots, x_i$. Entonces, por analogía con los espacios de segmentos dirigidos diremos que:

Se llama *volumen* $V(x_1, x_2, \ldots, x_n)$ *del sistema de vectores* $x_1, x_2, \ldots, x_n$ del espacio euclídeo E_n el valor que toma en este sistema una función real que depende de n argumentos vectoriales de E_n y está definida del modo siguiente:

$$V(x_1, x_2, \ldots, x_n) = \prod_{i=0}^{n-1} |\mathrm{ort}_{L_i} x_{i+1}|. \qquad (35.4)$$

Naturalmente, por ahora no podemos afirmar que el volumen de un sistema de vectores posee todas las propiedades que son propias precisamente al volumen para cualquier valor de n. Pero, para los espacios euclídeos de dimensiones 1, 2, 3, respectivamente, en virtud del isomorfismo euclídeo y las correlaciones (35.3), el volumen tiene, a ciencia cierta, las mismas propiedades que la longitud de un segmento, el área de un paralelogramo y el volumen de un paralelepípedo, respectivamente.

Veamos ahora un acceso algo más diferente al concepto de volumen de un sistema de vectores del espacio euclídeo E_n. Como ya se ha indicado, la atribución de determinados signos transforma la longitud, el área y el volumen en funciones algebraicas con ciertas propiedades comunes. Por esta razón podemos esperar que la analogía correspondiente tenga lugar también en un espacio euclídeo arbitrario. Teniendo presente precisamente esta analogía, daremos a conocer la siguiente definición:

Se llama *volumen orientado* $V^{\pm}(x_1, x_2, \ldots, x_n)$ *del sistema de vectores* $x_1, x_2, \ldots, x_n$ del espacio euclídeo E_n el valor que toma en este sistema una función real que depende de n argumentos vectoriales de E_n y posee las propiedades (34.4).

En esta definición tampoco todo está claro. Se desconoce si existe o no el volumen orientado para cualquier sistema de vectores en un espacio euclídeo arbitrario cuando $n \geq 4$. Si incluso existe, ¿será unívoco el modo en que lo definen las propiedades (34.4)? Por fin, ¿qué relación existe, en el caso general, entre el volumen y el volumen orientado? Por ahora podemos responder solamente a la última pregunta y sólo en el caso en que $n = 1, 2, 3$.

A veces tendremos que considerar el volumen y el volumen orientado en el espacio E_n para los sistemas que contienen menos que n vectores. Esto será el indicio de que en realidad nos interesamos no

con todo el espacio, sino con un subespacio suyo en el cual se toma el sistema dado. Correspondientemente, las propiedades (34.4) serán consideradas sólo respecto de los vectores del mismo subespacio. Puede surgir la necesidad de considerar el volumen y el volumen orientado para unos sistemas que contienen más que n vectores. De conformidad con la fórmula (35.4) y la propiedad **B** de (34.4), ambas funciones en tales sistemas han de ser iguales a cero.

En conclusión observemos que el empleo de dos diferentes conceptos relacionados con el volumen permitirá simplificar, en grado considerable, su investigación, puesto que un concepto refleja el aspecto geométrico del problema a resolver, mientras que el otro, el aspecto algebraico. Muy pronto se pondrá de manifiesto la relación existente entre ellos. Descubriremos, a continuación, que la importancia que proviene de la introducción de estos conceptos consiste, además, en que ellos generan cierto aparato matemático cuya significación no se limita al problema de un volumen.

Ejercicios.

1. Demuéstrese que en un espacio de segmentos dirigidos la longitud, el área y el volumen orientados se definen por las condiciones (34.4) de un modo unívoco.
2. ¿Será unívoco el modo en que se definen los mismos conceptos, si se excluye una de las condiciones (34.4)?
3. Demuéstrese que en todo espacio euclídeo $V(x_1, x_2) = |x_1|\cdot|x_2|$ cuando, y sólo cuando, los vectores x_1 y x_2 sean ortogonales.
4. Demuéstrese que en cualquier espacio euclídeo $V(x_1, x_2) = V(x_2, x_1)$.

§ 36. Propiedades geométricas y algebraicas del volumen

La investigación del concepto de volumen en el espacio euclídeo E_n la empezamos con el estudio de sus propiedades geométricas y algebraicas que provienen de la definición.

PROPIEDAD 1. *Siempre* $V(x_1, x_2, \ldots, x_n) \geqslant 0$. *La igualdad* $V(x_1, x_2, \ldots, x_n) = 0$ *tiene lugar cuando, y sólo cuando, el sistema de vectores* $x_1, x_2, \ldots, x_n$ *es linealmente dependiente.*

La primera parte de la afirmación se deduce obviamente de (35.4) y la demostración se necesita sólo para la segunda parte. Supongamos que el sistema $x_1, x_2, \ldots, x_n$ es linealmente dependiente. Si es que $x_1 = 0$, entonces, de la definición se desprende que el volumen es también igual a cero. En cambio, si $x_1 \neq 0$, entonces cierto vector x_{k+1} se expresa linealmente en términos de los vectores anteriores $x_1, \ldots, x_k$. Pero, en tal caso, $\text{ort}_{L_k} x_{k+1} = 0$ y el volumen de nuevo es igual a cero.

Supongamos ahora que el volumen es igual a cero. Por definición esto significa que es igual a cero uno de los factores en el segundo miembro de (35.4). Sea $i = k$ para este factor. Si $k = 0$, se tiene $x_1 = 0$. En el caso de que $k \neq 0$, la condición $\mathrm{ort}_{L_k} x_{k+1} = 0$ significa que el vector x_{k+1} pertenece a la cápsula lineal formada por los vectores $x_1, \ldots, x_k$, es decir, el sistema $x_1, \ldots, x_{k+1}$ es linealmente dependiente. En ambos casos todo el sistema de vectores $x_1, x_2, \ldots, x_n$ está también linealmente dependiente.

PROPIEDAD 5. *Para cualquier sistema de vectores $x_1, x_2, \ldots, x_n$ es válida la desigualdad de Hadamard*

$$V(x_1, x_2, \ldots, x_n) \leqslant \prod_{i=0}^{n-1} |x_{i+1}|, \qquad (36.1)$$

con la particularidad de que la igualdad tiene lugar cuando, y sólo cuando, el sistema $x_1, x_2, \ldots, x_n$ es ortogonal o contiene el vector nulo.

Debido a las propiedades que poseen una perpendicular y una proyección, es válida, a ciencia cierta, la desigualdad

$$|\mathrm{ort}_{L_i} x_{i+1}| \leqslant |x_{i+1}|, \qquad (36.2)$$

con la particularidad de que la desigualdad se convierte en una igualdad cuando, y sólo cuando, $x_{i+1} \perp L_i$, o bien, que es lo mismo, cuando el vector x_{i+1} es ortogonal a los vectores $x_1, x_2, \ldots, x_i$. Consideremos los productos de los miembros primero y segundo de las desigualdades del tipo (36.2) para todo i. Tenemos

$$\prod_{i=0}^{n-1} |\mathrm{ort}_{L_i} x_{i+1}| \leqslant \prod_{i=0}^{n-1} |x_{i+1}|.$$

Si todos los vectores del sistema $x_1, x_2, \ldots, x_n$ son no nulos, esta desigualdad se convierte en una igualdad cuando, y sólo cuando, el sistema es ortogonal. La presencia del vector nulo hace el caso trivial.

De la desigualdad de Hadamard se pueden deducir unos cuantos corolarios útiles. Supongamos que el sistema $x_1, x_2, \ldots, x_n$ está normalizado. En este caso, evidentemente,

$$V(x_1, x_2, \ldots, x_n) \leqslant 1.$$

Será clara también la afirmación siguiente. Si el sistema $x_1, x_2, \ldots, x_n$ está normalizado y su volumen es igual a la unidad, será ortonormalizado. El hecho de que el volumen de cualquier sistema normalizado no es superior a la unidad, es testimonio de que entre todos los sistemas normalizados el sistema ortonormalizado tiene el volumen máximo.

PROPIEDAD 6. *Para cualesquiera dos conjuntos ortogonales de vectores $x_1, x_2, \ldots, x_p$ e $y_1, y_2, \ldots, y_r$ se verifica la igualdad*

$$V(x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_r) = \\ = V(x_1, x_2, \ldots, x_p)\, V(y_1, y_2, \ldots, y_r).$$

Designemos mediante L_t una cápsula lineal formada por los primeros t vectores del sistema unido $x_1, \ldots, x_k, y_1, \ldots, y_r$ y mediante K_t, la cápsula lineal formada por los vectores $y_1, \ldots, y_t$. Por hipótesis, cada uno de los vectores del sistema $y_1, \ldots, y_r$ es ortogonal a todos los vectores del sistema $x_1, \ldots, x_k$. Por esta razón

$$L_{k+t} = L_k \oplus K_t$$

para cualquier valor de t desde 0 hasta r. Ahora, al tomar en consideración la igualdad (30.10), tenemos

$$V(x_1, \ldots, x_k, y_1, \ldots, y_r) = \left(\prod_{i=0}^{k-1} |\mathrm{ort}_{L_i} x_{i+1}|\right) \left(\prod_{i=0}^{r-1} |\mathrm{ort}_{L_{k+i}} y_{i+1}|\right) =$$

$$= \left(\prod_{i=0}^{k-1} |\mathrm{ort}_{L_i} x_{i+1}|\right) \left(\prod_{i=0}^{r-1} |\mathrm{ort}_{K_i} y_{i+1}|\right) = V(x_1, \ldots, x_k)\, V(y_1, \ldots, y_r).$$

Antes de pasar a las investigaciones ulteriores haremos una observación. El volumen de un sistema se expresa sólo en términos de las perpendiculares a las cápsulas lineales formadas por los vectores anteriores. Teniendo presente las propiedades de las perpendiculares, se puede deducir que el volumen del sistema no se alterará, si a un vector cualquiera se le añade una combinación lineal de los vectores anteriores. En particular, el volumen quedará intacto, si cualquier vector se sustituye por una perpendicular trazada desde este vector a cualquier cápsula lineal formada por los vectores anteriores.

PROPIEDAD 4. *El volumen del sistema de vectores no se altera cualquiera que sea la permutación de los vectores del sistema.*

Examinemos primero el caso en que dentro del sistema de vectores $x_1, \ldots, x_n$ encuentran dos vectores contiguos x_p, x_{p+1}. De conformidad con la observación recién mencionada, el volumen no se alterará, si los vectores x_p, x_{p+1} son sustituidos por los vectores $\mathrm{ort}_{L_{p-1}} x_p$, $\mathrm{ort}_{L_{p-1}} x_{p+1}$ y los vectores $x_{p+2}, \ldots, x_n$ por los vectores $\mathrm{ort}_{L_{p+1}} x_{p+2}, \ldots, \mathrm{ort}_{L_{n-1}} x_n$. En estas condiciones tres conjuntos de vectores

$$x_1, \ldots, x_{p-1},$$

$$\mathrm{ort}_{L_{p-1}} x_p,\ \mathrm{ort}_{L_{p-1}} x_{p+1}, \qquad \mathrm{ort}_{L_{p+1}} x_{p+2}, \ldots, \mathrm{ort}_{L_{n-1}} x_n$$

son ortogonales dos a dos y, en virtud de la propiedad 3, tendremos

$$V(x_1, \ldots, x_p, x_{p+1}, \ldots, x_n) = V(x_1, \ldots, x_{p-1}) \times$$

$$\times V(\mathrm{ort}_{L_{p-1}} x_p, \mathrm{ort}_{L_{p-1}} x_{p+1}) \cdot V(\mathrm{ort}_{L_{p+1}} x_{p+2}, \ldots, \mathrm{ort}_{L_{n-1}} x_n).$$

Está claro que las cápsulas lineales de los vectores $x_1, \ldots, x_p, x_{p+1}$ y $x_1, \ldots, x_{p-1}, x_{p+1}, x_p$ coinciden y, por consiguiente,

$$V(x_1, \ldots, x_{p+1}, x_p, \ldots, x_n) = V(x_1, \ldots, x_{p-1}) \times$$

$$\times V(\mathrm{ort}_{L_{p-1}} x_{p+1}, \mathrm{ort}_{L_{p-1}} x_p) \cdot V(\mathrm{ort}_{L_{p+1}} x_{p+2}, \ldots, \mathrm{ort}_{L_{n-1}} x_n).$$

En vista del isomorfismo euclídeo, el volumen de un sistema de dos vectores posee las mismas propiedades que el área de un paralelogramo. En particular, no depende del orden de los vectores del sistema. Comparando los primeros miembros de dos últimas igualdades, concluimos que

$$V(x_1, \ldots, x_{p+1}, x_p, \ldots, x_n) = V(x_1, \ldots, x_p, x_{p+1}, \ldots, x_n).$$

Un poco más tarde demostraremos que cualquier permutación $x_{j_1}, x_{j_2}, \ldots, x_{j_n}$ de vectores del sistema $x_1, x_2, \ldots, x_n$ puede ser obtenida por la permutación consecutiva de los vectores contiguos. Por ello, la propiedad 4 para una permutación arbitraria se desprende del caso particular examinado.

PROPIEDAD 5. *El volumen de un sistema de vectores es una función absolutamente homogénea, es decir,*

$$V(x_1, \ldots, \alpha x_p, \ldots, x_n) = |\alpha| \, V(x_1, \ldots, x_p, \ldots, x_n)$$

para cualquier valor de p.

De acuerdo con la propiedad 4, no disminuiremos la generalidad, si consideramos que $p = n$. Pero, en tal caso, al tomar en consideración (30.6), obtendremos

$$V(x_1, \ldots, x_{n-1}, \alpha x_n) = \left(\prod_{i=0}^{n-2} |\mathrm{ort}_{L_i} x_{i+1}|\right) |\mathrm{ort}_{L_{n-1}} (\alpha x_n)| =$$

$$= |\alpha| \prod_{i=0}^{n-1} |\mathrm{ort}_{L_i} x_{i+1}| = |\alpha| \, V(x_1, \ldots, x_{n-1}, x_n).$$

PROPIEDAD 6. *El volumen de un sistema de vectores no cambia, si a uno de los vectores del sistema se le añade una combinación lineal de todos los demás vectores.*

Recurriendo de nuevo a la propiedad 4, podemos considerar que al último vector se le agrega una combinación lineal de los vectores anteriores. Pero, según se ha observado, en el caso dado el volumen no cambia.

El volumen de un sistema de vectores es una función real. Esta función posee ciertas propiedades, una parte de las cuales ya la hemos establecido. Se ha confirmado nuestra suposición de que el volumen del sistema de vectores, determinado por nosotros para el espacio euclídeo, posee todas las propiedades propias precisamente del volumen para cualquier valor de n. Pero, lo más importante consiste, quizás, en que las propiedades establecidas definen el volumen de modo unívoco. Más preciso es válido el

TEOREMA 36.1. *Si una función real $F(x_1, x_2, \ldots, x_n)$, que depende de n argumentos vectoriales de E_n, posee las siguientes propiedades:*

A) *no cambia, si a cualquier argumento se le añade cualquiera combinación lineal de los argumentos restantes,*

B) *es absolutamente homogénea,* (36.3)

C) *es igual a la unidad para todos los sistemas ortonormalizados, entonces coincide con el volumen del sistema de vectores.*

DEMOSTRACIÓN. Si entre los argumentos $x_1, \ldots, x_n$ se tiene al menos uno que es nulo, entonces, de conformidad con la propiedad B,

$$F(x_1, x_2, \ldots, x_n) = V(x_1, x_2, \ldots, x_n) = 0. \qquad (36.4)$$

Supongamos ahora que el sistema $x_1, x_2, \ldots, x_n$ es arbitrario. Al sustituir de todo vector x_i su proyección sobre un subespacio formado por los vectores $x_1, \ldots, x_{i-1}$ y teniendo presente la propiedad A, concluimos que

$$F(x_1, x_2, \ldots, x_n) = F(\mathrm{ort}_{L_0} x_1, \mathrm{ort}_{L_1} x_2, \ldots, \mathrm{ort}_{L_{n-1}} x_n). \qquad (36.5)$$

Si el sistema $x_1, x_2, \ldots, x_n$ es linealmente dependiente, entonces entre los vectores $\mathrm{ort}_{L_{i-1}} x_i$ se tiene por lo menos un vector nulo y la igualdad (36.4) también tiene lugar. Supongamos que el sistema $x_1, x_2, \ldots, x_n$ es linealmente independiente; en este caso todos los vectores del sistema

$$\mathrm{ort}_{L_0} x_1, \mathrm{ort}_{L_1} x_2, \ldots, \mathrm{ort}_{L_{n-1}} x_n$$

serán no nulos. Como que dicho sistema es, además, ortogonal, existe un sistema ortonormalizado $e_1, e_2, \ldots, e_n$ para el cual

$$\mathrm{ort}_{L_{i-1}} x_i = |\mathrm{ort}_{L_{i-1}} x_i| \, e_i.$$

Debido a la propiedad C, $F(e_1, e_2, \ldots, e_n) = 1$.

Por eso, de la propiedad B se desprende que

$$F(x_1, x_2, \ldots, x_n) = \left(\prod_{i=0}^{n-1} |\mathrm{ort}_{L_i} x_{i+1}|\right) F(e_1, e_2, \ldots, e_n) = V(x_1, x_2, \ldots, x_n).$$

El teorema demostrado permite afirmar que si construimos de tal o cual modo una función que posee las propiedades (36.3), entonces esta función será el volumen del sistema de vectores.

Ejercicios.

1. Establézcase el sentido geométrico de las propiedades 2, 3, 6 en los espacios de segmentos dirigidos.
2. Establézcase el sentido geométrico de la igualdad (36.5) en los espacios de segmentos dirigidos.
3. ¿Podrá una función, que satisface las condiciones (36.3), ser nula en un sistema de vectores linealmente independiente?
4. Supongamos que respecto de una base ortonormalizada $e_1, e_2, \ldots, e_n$ el sistema de vectores $x_1, x_2, \ldots, x_n$ posee la propiedad

$$(x_i, e_j) = 0$$

para $i = 2, 3, \ldots, n$ y $j < i$ ($i = 1, 2, \ldots, n-1$ y $j > i$). Hállese la expresión para el volumen $V(x_1, x_2, \ldots, x_n)$ en términos de las coordenadas de los vectores $x_1, x_2, \ldots, x_n$ en la base $e_1, e_2, \ldots, e_n$.

5. ¿Qué cambiará, si consideramos el concepto de volumen en un espacio complejo?

§ 37. Propiedades algebraicas del volumen orientado

Pasemos ahora a la investigación de las propiedades algebraicas del volumen orientado, dejando de lado hasta el momento la cuestión de su existencia. Tomemos como base de las investigaciones las condiciones A, B, C de (34.4).

PROPIEDAD 1. *El volumen orientado de un sistema de vectores es nulo, si dos vectores cualesquiera coinciden.*

Esta propiedad es un corolario directo de la condición B. Es fácil demostrar que, cumpliéndose la condición A, la condición B y la propiedad enunciada 1 son equivalentes.

PROPIEDAD 2. *El volumen orientado de un sistema de vectores cambia su signo, al cambiar de lugar dos vectores cualesquiera.*

La demostración es análoga para los vectores de toda clase, por lo que con el fin de simplificar las anotaciones nos limitaremos al examen del caso cuando cambiamos de lugar los vectores primero y segundo. De acuerdo con la propiedad 1,

$$V_{\pm}(x_1 + x_2,\ x_1 + x_2,\ x_3, \ldots, x_n) = 0.$$

Pero, por otro lado, según la condición A

$$V_{\pm}(x_1 + x_2, x_1 + x_2, x_3, \ldots, x_n) = V_{\pm}(x_1, x_1, x_3, \ldots, x_n) + V_{\pm}(x_2, x_2, x_3, \ldots, x_n) + V_{\pm}(x_1, x_2, x_3, \ldots, x_n) + V_{\pm}(x_2, x_1, x_3, \ldots, x_n).$$

En el segundo miembro de esta igualdad los primeros dos sumandos son nulos, de donde se deduce la validez de la propiedad 2. Este tampoco es difícil demostrar que si se cumple la condición A, la condición B y la propiedad enunciada 2 son equivalentes.

PROPIEDAD 3. *El volumen orientado de un sistema de vectores queda inalterable, si a cierto vector se le agrega una combinación cualquiera de los vectores restantes.*

Para simplificar, consideremos sólo el primer vector. Según la condición A tenemos $V_{\pm}(x_1 + \sum_{i=2}^{n} \alpha_i x_i, x_2, \ldots, x_n) =$

$$= V_{\pm}(x_1, x_2, \ldots, x_n) + \sum_{i=2}^{n} \alpha_i V_{\pm}(x_i, x_2, \ldots, x_n).$$

En esta igualdad todos los sumandos del segundo miembro, a excepción del primero, son nulos, de acuerdo con la propiedad 1.

PROPIEDAD 4. *El volumen orientado es una función homogénea, es decir, para cualquier valor de p se verifica la correlación*

$$V_{\pm}(x_1, \ldots, \alpha x_p, \ldots, x_n) = \alpha V_{\pm}(x_1, \ldots, x_p, \ldots, x_n).$$

Esta propiedad es un corolario inmediato de la condición A.

PROPIEDAD 5. *La igualdad $V_{\pm}(x_1, x_2, \ldots, x_n) = 0$ tiene lugar cuando, y sólo cuando, el sistema de vectores $x_1, x_2, \ldots, x_n$ es linealmente dependiente.*

Evidentemente, es preciso sólo demostrar que de la igualdad $V_{\pm}(x_1, x_2, \ldots, x_n) = 0$ se desprende la dependencia lineal de los vectores $x_1, x_2, \ldots, x_n$. Supondremos lo contrario. Admitamos que el volumen orientado es igual a cero para cierto sistema linealmente independiente $y_1, y_2, \ldots, y_n$. Este sistema es una base de E_n, razón por la cual para todo vector z de E_n se tiene

$$z = \alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_n y_n.$$

Sustituyamos en el sistema $y_1, y_2, \ldots, y_n$ un vector cualquiera, por ejemplo, y_1 por el vector z. Haciendo uso sucesivamente de las propiedades 3 y 4, encontramos que

$$V_{\pm}(z, y_2, \ldots, y_n) = V_{\pm}\left(\alpha_1 y_1 + \sum_{i=2}^{n} \alpha_i y_i, y_2, \ldots, y_n\right) =$$
$$= V_{\pm}(\alpha_1 y_1, y_2, \ldots, y_n) = \alpha_1 V_{\pm}(y_1, y_2, \ldots, y_n) = 0.$$

Por definición, el volumen orientado no es nulo al menos en un sistema linealmente independiente $z_1, z_2, \ldots, z_n$. Pero, al sustituir sucesivamente los vectores $y_1, y_2, \ldots, y_n$ por los vectores $z_1, z_2, \ldots, z_n$, llegamos a que, de conformidad con el teorema 15.2, en este sistema el volumen orientado también es igual a cero. La contradicción obtenida demuestra la propiedad en consideración.

PROPIEDAD 6. *Si dos volúmenes orientados coinciden al menos en un sistema linealmente independiente de vectores, su coincidencia es idéntica.*

Supongamos que para nosotros es conocido que los volúmenes orientados $V_1^{\pm}(x_1, x_2, \ldots, x_n)$ y $V_2^{\pm}(x_1, x_2, \ldots, x_n)$ coinciden en el sistema linealmente independiente $z_1, z_2, \ldots, z_n$. Veamos la diferencia $F(x_1, x_2, \ldots, x_n) = V_1^{\pm}(x_1, x_2, \ldots, x_n) - V_2^{\pm} \times \times (x_1, x_2, \ldots, x_n)$. Esta función satisface las propiedades 3 y 4 del volumen orientado. Además, es nula en todos los sistemas linealmente dependientes y por lo menos en un sistema linealmente independiente $z_1, z_2, \ldots, z_n$. Repitiendo los razonamientos empleados en la demostración de la propiedad 5, concluimos que $F(x_1, x_2, \ldots$

..., x_n) es igual a cero en todos los sistemas linealmente independientes, es decir, que es nulo idénticamente.

De la propiedad 6 se infiere que el volumen orientado se define por las condiciones (34.4) de modo único, si se fija aquel sistema ortonormalizado en que el volumen debe ser igual a la unidad.

PROPIEDAD 7. *El módulo del volumen orientado de un sistema de vectores coincide con el volumen del mismo sistema.*

Supongamos que el volumen orientado es igual a la unidad en el sistema ortonormalizado $z_1, z_2, \ldots, z_n$. Examinemos las funciones $|V_{\pm}(x_1, x_2, \ldots, x_n)|$ y $V(x_1, x_2, \ldots, x_n)$. Ambas satisfacen las condiciones A y B de (36.3) y coinciden en el sistema linealmente independiente $z_1, z_2, \ldots, z_n$. La función

$$\Phi(x_1, x_2, \ldots, x_n) = ||V_{\pm}(x_1, x_2, \ldots, x_n)| - V(x_1, x_2, \ldots, x_n)|$$

satisface también las condiciones A, B, de (36.3), es nula en todos los sistemas linealmente dependientes y por lo menos en un sistema linealmente independiente $z_1, z_2, \ldots, z_n$. Al repetir de nuevo los razonamientos realizados en la demostración de la propiedad 5, concluimos que $\Phi(x_1, x_2, \ldots, x_n)$ es nula idénticamente.

La última propiedad es de mucha importancia, puesto que permite afirmar que el módulo del volumen orientado debe poseer las mismas propiedades que tiene el volumen. En particular, debe ser igual en magnitud absoluta a la unidad en todos los sistemas ortonormalizados y no sólo en uno de ellos. Para el módulo es verídica la desigualdad de Hadamard, etc. Esta propiedad nos da la respuesta definitiva a todas las preguntas planteadas respecto del volumen y del volumen orientado. Lo único que nos falta es la demostración de existencia del volumen orientado.

Ejercicios.

1. Demuéstrese que si se cumple la condición A de (34.4), la condición B es equivalente tanto a la propiedad 1, como a la 2.
2. Demuéstrese que, cualquiera que sea el número real α, existe un sistema de vectores, para el cual el volumen orientado es igual a α.
3. Supongamos que la condición C de (34.4) ha sido sustituida por la condición de igualdad a cualquier número fijado en todo sistema linealmente independiente. ¿Cómo cambiará el volumen orientado?
4. ¿Se usaba en la deducción de las propiedades del volumen orientado el hecho de que en el espacio lineal existe el producto escalar y que es real el volumen orientado? ¿Qué cambiará, si consideramos el volumen orientado en un espacio complejo?

§ 38. Permutaciones

Examinaremos un sistema $x_1, x_2, \ldots, x_n$ y otro sistema, $x_{j_1}, x_{j_2}, \ldots, x_{j_n}$, obtenido del primero realizando diversas permutaciones de vectores. Supongamos que de un sistema puede obtenerse el otro realizando permutaciones

consecutivas sólo entre los pares de elementos de cada sistema. En tal caso los volúmenes de dichos sistemas serán iguales y los volúmenes orientados, o bien iguales o bien se diferenciarán en el signo, lo cual depende de la cantidad de permutaciones que han de realizarse.

En cuestiones de interés sobre permutaciones, las propiedades individuales de los vectores no desempeñarán ningún papel, pero será importante el orden de éstos. Por eso, en lugar de los propios vectores consideraremos los números 1, 2, . . ., n. La totalidad de números

$$i_1, i_2, \ldots, i_n,$$

entre los cuales no hay números iguales y cada uno de los cuales es uno de los números 1, 2, . . ., n se denomina *permutación* de estos números. La permutación 1, 2, . . ., n se denomina *normal*.

Es fácil señalar que en un conjunto de n números la cantidad total de toda clase de permutaciones es igual a $n!$. En efecto, para $n = 1$ esto es evidente. Supongamos que la afirmación es justa para cualquier conjunto de $n - 1$ números. Todas las permutaciones de n números se pueden dividir en n clases, incluyendo en una clase sólo aquellas permutaciones cuyo primer número es el mismo. La cantidad de permutaciones en cada clase coincide con la cantidad de permutaciones de $n - 1$ números, es decir, es igual a $(n - 1)!$. Por consiguiente, la cantidad de todas las permutaciones de n números es igual a $n!$.

Suele decirse que en una permutación dada los números i, j forman una *inversión*, si $i > j$, pero i precede a j en la permutación. Una permutación se llamará *par*, siempre que sus números constituyan una cantidad par de inversiones y se llamará *impar*, en el caso contrario. Si en una permutación se cambian de lugar dos números cualesquiera, no necesariamente contiguos, y todos los demás números quedan en sus lugares, obtendremos una permutación nueva. Esta transformación de la permutación se denomina *transposición*.

Demostremos que toda transposición altera la paridad de la permutación. Para los números contiguos esta afirmación es obvia. Su disposición mutua respecto de los demás números quedó la misma, mientras que la permutación de los propios números cambió la cantidad total de inversiones en una unidad.

Supongamos ahora que entre los números a permutar i y j se hallan otros s números $k_1, k_2, \ldots, k_s$, es decir, la permutación tiene por expresión

$$\ldots, i, k_1, k_2, \ldots, k_s, j, \ldots$$

Cambiaremos sucesivamente de lugar el número i con los números contiguos $k_1, k_2, \ldots, k_s, j$. A continuación, el número j, que

ya precede a i, lo trasladamos a la izquierda con ayuda de s transposiciones con los números $k_s, k_{s-1}, \ldots, k_1$. De este modo, en total llevaremos a cabo $2s+1$ transposiciones de los números que están junto uno al otro. Por consiguiente, la paridad de la transposición cambiará.

TEOREMA 33.1. *Todas las $n!$ permutaciones de n números se pueden disponer en un orden tal que cada permutación siguiente se obtenga de la anterior con ayuda de una transposición, con la particularidad de que para el inicio del proceso sirve cualquier permutación.*

DEMOSTRACIÓN. Esta afirmación es justa para $n = 2$. Si es necesario empezar por la permutación 1, 2, entonces la disposición buscada será 1, 2, 2, 1; en cambio, si empezamos con la permutación 2, 1, entonces la disposición buscada será 2, 1, 1, 2.

Supongamos que el teorema ya se ha demostrado para cualesquiera permutaciones que contienen no más de $n-1$ números. Consideremos las permutaciones de n números. Supongamos que debemos empezar por la permutación $i_1, i_2, \ldots, i_n$. La disposición de las permutaciones se realizará según la siguiente regla. Comencemos por las permutaciones que en el primer lugar tienen el número i_1. Por hipótesis, se puede ordenar todas estas permutaciones en concordancia con los requisitos del teorema, puesto que, de hecho, es preciso disponer en el orden necesario todas las permutaciones de $n-1$ números.

En la última permutación, obtenida de esta manera, realizamos una transposición, trasladando al primer lugar el número i_2. A continuación, ordenamos, igual que en el caso precedente, todas las permutaciones que en el primer lugar tienen este número dado, etc. Por el procedimiento indicado podemos probar todas las permutaciones de n números.

Con tal sistema de disposición de las permutaciones de n números las permutaciones contiguas tendrán paridades contrarias. Tomando en consideración la paridad del número $n!$ para $n \geqslant 2$, podemos concluir que en este caso la cantidad de permutaciones pares de n números es igual a la cantidad de permutaciones impares y equivale a $\frac{1}{2} n!$.

Ejercicios.

1. ¿Cuál es la paridad de la permutación 5, 2, 3, 1, 4?
2. Demuéstrese que ninguna de las permutaciones pares (impares) puede ser reducida a la normal en el transcurso de un número impar (par) de transposiciones.
3. Examinemos un par de permutaciones $i_1, i_2, \ldots, i_n$ y $1, 2, \ldots, n$. Reduciremos a la forma normal la primera permutación con ayuda de transposiciones, realizando en cada una de éstas una transposición de cualesquiera elementos en la segunda permutación. Demuéstrese que terminando el proceso, la segunda permutación tendrá la misma paridad que la permutación $i_1, i_2, \ldots, i_n$.

§ 39. Existencia de un volumen orientado

Examinemos ahora la cuestión sobre la existencia de un volumen orientado de un sistema de vectores. Supongamos que en el espacio E_n se ha elegido un sistema ortonormalizado $e_1, e_2, \ldots, e_n$ en el cual el volumen orientado ha de ser igual a uno, de acuerdo con la condición C de (34.4). Tomemos un sistema arbitrario $z_1, z_2, \ldots, z_n$ de vectores de E_n. Puesto que el sistema $e_1, e_2, \ldots, e_n$ es la base en E_n, para todo vector z_i existe la siguiente descomposición según la base citada:

$$z_i = a_{i1}e_1 + a_{i2}e_2 + \ldots + a_{in}e_n, \tag{39.1}$$

donde a_{ij} son ciertos números.

Si el volumen orientado existe, de conformidad con la condición A de (34.4), podemos transformarlo sucesivamente teniendo presente las descomposiciones (39.1). A saber,

$$V_{\pm}(z_1, z_2, \ldots, z_n) = V_{\pm}\Big(\sum_{j_1=1}^{n} a_{1j_1}e_{j_1}, \sum_{j_2=1}^{n} a_{2j_2}\times$$

$$\times e_{j_2}, \ldots, \sum_{j_n=1}^{n} a_{nj_n}e_{j_n}\Big) =$$

$$= \sum_{j_1=1}^{n} a_{1j_1}V_{\pm}\Big(e_{j_1}, \sum_{j_2=1}^{n} a_{2j_2}e_{j_2}, \ldots, \sum_{j_n=1}^{n} a_{nj_n}e_{j_n}\Big) =$$

$$= \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_{1j_1}a_{2j_2}V_{\pm}\Big(e_{j_1}, e_{j_2}, \ldots, \sum_{j_n=1}^{n} a_{nj_n}e_{j_n}\Big) = \ldots$$

$$\ldots = \sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\ldots\sum_{j_n=1}^{n} a_{1j_1}a_{2j_2}\ldots a_{nj_n}V_{\pm}(e_{j_1}, e_{j_2}, \ldots, e_{j_n}). \tag{39.2}$$

En la última suma múltiple de n la mayor parte de los sumandos es igual a cero, puesto que, de acuerdo con la propiedad 1, el volumen orientado del sistema de vectores es nulo, siempre que dos vectores cualesquiera coincidan. Por ello, entre los sistemas $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ se deben considerar sólo aquellos, para los cuales el surtido de índices $j_1, j_2, \ldots, j_n$ representa una permutación de n números $1, 2, \ldots, n$. Mas, en este caso

$$V_{\pm}(e_{j_1}, e_{j_2}, \ldots, e_{j_n}) = \pm 1$$

según sea par o impar la permutación de los índices.

De este modo, si el volumen orientado existe, debe expresarse en términos de las coordenadas de los vectores $z_1, z_2, \ldots, z_n$

en la base $z_1, z_2, \dots, z_n$ mediante la siguiente fórmula:

$$V^{\pm}(x_1, x_2, \dots, x_n) = \sum \pm a_{1j_1} a_{2j_2} \dots a_{nj_n}. \qquad (39.3)$$

Aquí la sumación se realiza según todas las permutaciones de índices $j_1, j_2, \dots, j_n$ de los números $1, 2, \dots, n$ y el signo más o el signo menos se toma en dependencia de la paridad o imparidad de la permutación.

Demostremos que una función, prefijada por el segundo miembro de la igualdad (39.3), satisface todas las condiciones que definen un volumen orientado. Sea el vector x_p una combinación lineal de los vectores x'_p y x''_p, es decir,

$$x_p = \alpha x'_p + \beta x''_p$$

para ciertos números α, β. Designemos por a'_{pj} y a''_{pj} las coordenadas de los vectores x'_p y x''_p, respectivamente, en la base $z_1, z_2, \dots, z_n$. Entonces, es evidente que

$$a_{pj} = \alpha a'_{pj} + \beta a''_{pj}$$

para todo j desde 1 hasta n. Luego hallamos

$$\sum \pm a_{1j_1} \dots a_{pj_p} \dots a_{nj_n} = \sum \pm a_{1j_1} \dots$$
$$\dots (\alpha a'_{pj_p} + \beta a''_{pj_p}) \dots a_{nj_n} = \alpha \sum \pm a_{1j_1} \dots$$
$$\dots a'_{pj_p} \dots a_{nj_n} + \beta \sum \pm a_{1j_1} \dots a''_{pj_p} \dots a_{nj_n},$$

y la condición A de (34.4) queda cumplida.

Supongamos que permutamos dos vectores cualesquiera del sistema $x_1, x_2, \dots, x_n$. En este caso la función (39.3) cambiará de signo, puesto que variará la paridad de cada permutación. Como ya se ha indicado antes, si es válida la propiedad de linealidad respecto de cada argumento, la propiedad demostrada equivaldrá al cumplimiento de la condición B de (34.4).

Y, por fin, examinemos el significado de la función construida sobre el sistema de vectores $z_1, z_2, \dots, z_n$. Para este sistema, las coordenadas a_{ij} tienen la forma

$$a_{ij} = \begin{cases} 0, & \text{si } i \neq j, \\ 1, & \text{si } i = j. \end{cases}$$

Por consiguiente, entre los sumandos de (39.3) será no nulo solamente el sumando $a_{11}a_{22} \dots a_{nn}$. La permutación $1, 2, \dots, n$ es par, los elementos $a_{11}, a_{22}, \dots, a_{nn}$ son iguales a uno, por lo cual el valor de la función en el sistema ortonormalizado $z_1, z_2, \dots, z_n$ es igual a la unidad.

Así pues, todas las condiciones (34.4) quedan cumplidas y la función (39.3) representa la expresión del volumen orientado del

sistema de vectores en los términos de las coordenadas. Esta expresión es *única* en virtud de la unicidad del volumen orientado.

Ejercicios.

1. ¿Se ha usado en esencia el carácter ortonormalizado del sistema $e_1, e_2, \dots, e_n$ en la deducción de la fórmula (39.3)?
2. ¿Qué cambios tendrán lugar en la fórmula (39.3), si en la condición C de (34.4) el volumen orientado se considera distinto de uno?
3. ¿Cómo esencial fue el uso de la condición B de (34.4) al deducir la fórmula (39.3)?
4. ¿Cambiará la fórmula (39.3) su forma, si el volumen orientado se considera en un espacio complejo?

§ 40. Determinantes

Supongamos que los vectores $x_1, x_2, \dots, x_n$ del espacio euclídeo R_n están dados por sus coordenadas

$$x_i = (a_{i1}, a_{i2}, \dots, a_{in})$$

en la base (21.7). Dispondremos los números a_{ij} en forma de una tabla A del modo siguiente:

$$A = \begin{pmatrix} a_{11} a_{12} \dots a_{1n} \\ a_{21} a_{22} \dots a_{2n} \\ \dots\dots\dots \\ a_{n1} a_{n2} \dots a_{nn} \end{pmatrix}$$

Esta tabla se denomina *matriz cuadrada de orden n*, los números a_{ij} son *elementos* de la matriz. Si numeramos las filas de una matriz consecutivamente de arriba abajo y las columnas, de izquierda a derecha, entonces el primer índice de un elemento significa el número de la fila en que se halla el elemento, y el segundo índice, número de la columna. Con respecto a los elementos $a_{11}, a_{22}, \dots, a_{nn}$ se dice que ellos forman la *diagonal principal* de la matriz A.

Cualesquiera n^2 números pueden disponerse en forma de una matriz cuadrada de orden n. Si los elementos de la fila de una matriz se consideran como coordenadas de cierto vector de R_n en la base (21.7), entonces entre todas las matrices cuadradas de orden n y los sistemas ordenados de n vectores del espacio R_n se establece una correspondencia biunívoca.

En el espacio R_n, al igual que en cualquier otro espacio, existe un volumen orientado. Será, además, el único, si exigimos que se cumpla la condición C de (34.4) en el sistema de vectores (21.7). Al tomar en consideración la citada correspondencia biunívoca, concluimos que en el conjunto de todas las matrices cuadradas se genera una función bien determinada. Teniendo presente (39.3), llegamos a la siguiente definición de esta función.

Se llama *determinante* de orden n, correspondiente a la matriz A, la suma algebraica de $n!$ términos compuesta de la siguiente manera. Como términos del determinante intervienen toda una serie de productos de n elementos de la matriz, uno en cada fila y en cada columna. El término se toma con signo más, si los índices de las columnas de sus elementos forman una permutación par, a condición de que los propios elementos están dispuestos en orden creciente de los números de las filas; el signo menos se toma en el caso contrario. Para designar un determinante usaremos el símbolo siguiente:

$$\det \begin{pmatrix} a_{11} a_{12} \ldots a_{1n} \\ a_{21} a_{22} \ldots a_{2n} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{n1} a_{n2} \ldots a_{nn} \end{pmatrix}$$

si es preciso indicar la forma explícita de los elementos de la matriz. En cambio, si esto no es necesario, se usará un símbolo más sencillo $\det A$, donde se indica sólo la designación de la matriz A. Los elementos de la matriz del determinante se llamarán también *elementos del determinante*.

El determinante coincide con el volumen orientado del sistema de filas de la matriz. Por esta razón, al investigarlo se pueden utilizar todos los datos conocidos referentes al volumen y al volumen orientado. En particular, un determinante es igual a cero cuando, y sólo cuando, las filas de la matriz son linealmente dependientes; el determinante cambia de signo cuando se intercambian de lugar dos filas cualesquiera, etc. Ahora nuestras investigaciones tocarán aquellas propiedades del determinante que son difíciles de demostrar sin emplear la expresión explícita del determinante en términos de los elementos de la matriz.

Llamemos *transposición* de una matriz la transformación de esta última en que las filas se convierten en columnas y las columnas, en filas con los mismos números. Una matriz, transpuesta respecto a la matriz A, se indica con A'. Suele decirse en este caso que el determinante

$$\det \begin{pmatrix} a_{11} a_{21} \ldots a_{n1} \\ a_{12} a_{22} \ldots a_{n2} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{1n} a_{2n} \ldots a_{nn} \end{pmatrix}$$

o $\det A'$ se ha obtenido por transposición del determinante (40.1). Con relación a la transposición, un determinante posee las siguientes propiedades de importancia.

Un determinante de cualquier matriz no varía durante la operación de transposición.

En efecto, el determinante de la matriz A está compuesto por

los términos del tipo

$$a_{1j_1}a_{2j_2}\dots a_{nj_n}, \qquad (40.2)$$

cuyo signo se determina por la paridad de la permutación $j_1, j_2, \dots, j_n$. Todos los factores del producto (40.2) en la matriz transpuesta A' quedan en distintas filas y distintas columnas, es decir, su producto es un término del determinante transpuesto. Designemos los elementos de la matriz A' mediante a'_{ij}. Está claro que $a'_{ij} = a_{ji}$, por lo cual

$$a_{1j_1}a_{2j_2}\dots a_{nj_n} = a'_{j_1 1}a'_{j_2 2}\dots a'_{j_n n}. \qquad (40.3)$$

Ordenemos los elementos del segundo miembro de (40.3) según el orden de crecimiento de los números de filas. En este caso la permutación de los índices de columnas tendrá la misma paridad que la permutación $j_1, j_2, \dots, j_n$. Mas, esto significa que el signo del término (40.2) en el determinante transpuesto será el mismo que en el determinante inicial. Por consiguiente, ambos determinantes se componen de unos mismos términos de signos iguales, es decir, ellos coinciden.

De la propiedad demostrada proviene que en un determinante las filas y columnas se encuentran en las mismas condiciones. Por esto, todas las propiedades demostradas anteriormente respecto a las filas serán justas también para las columnas.

Consideremos un determinante d de orden n. Elijamos arbitrariamente en su matriz k filas y k columnas. Los elementos que se hallan en la intersección de las filas y columnas elegidas forman una matriz de orden k. El determinante de esta matriz se denomina *menor* de orden k del determinante d. El menor dispuesto en las primeras k columnas y las primeras k filas se llama menor *principal* o menor *angular*.

Supongamos ahora que en el determinante d de orden n se ha elegido un menor M de orden k. Si suprimimos aquellas filas y columnas en cuyas intersecciones se halla el menor M, quedará el menor N de orden $n-k$. Este se llamará menor *complementario* de M. En cambio, si suprimimos las filas y columnas en las cuales se hallan los elementos del menor N, quedará, evidentemente, el menor M. De este modo, se puede hablar de un par de menores recíprocamente complementarios.

Si el menor M de orden k se halla en las filas de números $i_1, i_2, \dots, i_k$ y en las columnas de números $j_1, j_2, \dots, j_k$, entonces el número

$$(-1)^{\sum_{p=1}^{k}(i_p+j_p)}N \qquad (40.4)$$

se llamará *complemento algebraico* del menor M.

TEOREMA 40.1. (*teorema de Laplace*). *Supongamos que en el determinante d de orden n se han elegido arbitrariamente k filas (columnas), donde $1 \leqslant k \leqslant n-1$. Entonces, la suma de los productos de todos los menores de k-ésimo orden, contenidos en las filas (columnas) elegidas, por sus complementos algebraicos es igual al determinante d.*

DEMOSTRACIÓN. Consideremos las columnas de la matriz del determinante d como vectores $x_1, x_2, \ldots, x_n$ del espacio R_n. La suma de los productos de todos los menores de k-ésimo orden, contenidos en las filas elegidas, por sus complementos algebraicos puede considerarse como cierta función $F(x_1, x_2, \ldots, x_n)$, dependiente de los vectores $x_1, x_2, \ldots, x_n$.

Esta función es, a ciencia cierta, lineal respecto de todo argumento, ya que la propiedad dada la poseen tanto los menores como los complementos algebraicos. Dicha función es igual a uno en el sistema ortonormalizado (21.7), de lo que es fácil convencerse por comprobación inmediata. Si demostramos que $F(x_1, x_2, \ldots, x_n)$ cambia de signo, al cambiar de lugar dos vectores cualesquiera, la función coincidirá con el volumen orientado del sistema de vectores $x_1, x_2, \ldots, x_n$. Pero el volumen orientado coincide con el determinante de la matriz en la que las coordenadas de los vectores están dispuestas en las filas. Puesto que el determinante de la matriz coincide con el de la matriz transpuesta, la demostración del teorema de Laplace se dará por terminado.

Evidentemente, sólo es necesario considerar un caso en que conmutan dos vectores contiguos, pues la permutación de cualesquiera dos vectores siempre se reduce a un número impar de permutaciones de vectores contiguos. La demostración de esta afirmación se ha aducido en el párrafo 38.

Supongamos que conmutan los vectores x_i y x_{i+1}. Establezcamos una correspondencia biunívoca entre los menores en las filas elegidas del determinante inicial y del determinante con las columnas permutadas. Designemos mediante ω la totalidad de los números de las columnas que definen el menor. Son posibles los siguientes casos:

1) $i, i+1 \in \omega$,

2) $i, i+1 \overline{\in} \omega$,

3) $i \in \omega, i+1 \overline{\in} \omega$,

4) $i+1 \in \omega, i \overline{\in} \omega$.

En los casos 1, 2 a todo menor le pondremos en correspondencia otro menor, dispuesto en las columnas con la totalidad de números ω; en los casos 3, 4, un menor, dispuesto en las columnas con

la totalidad de números obtenida de α mediante la sustitución de i por $i+1$ e $i+1$ por i, respectivamente.

Hemos de notar que en todos los casos los menores correspondientes se determinan mediante una misma totalidad de elementos. Más aún, en los casos 2—4 los menores coinciden y en el caso 1, sólo difieren en el signo, puesto que en uno de ellos están permutadas, respecto del otro, dos columnas. En virtud de las razones análogas, en el caso 2 los menores complementarios correspondientes se diferencian en el signo, siendo coincidentes en los casos restantes. Los complementos algebraicos se diferencian de los menores complementarios sólo en el signo, determinado por la paridad de la suma de los números de las filas y columnas en las cuales se encuentra el menor. En los casos 1, 2 estos números de los menores correspondientes son idénticos y en los casos 3, 4 difieren en unidad.

Comparando ahora los sumandos correspondientes de la función $F(x_1, x_2, \ldots, x_n)$ y de una función obtenida por permutación de los vectores x_i y x_{i+1}, observamos que difieren sólo en el signo. Por consiguiente, en la permutación de dos vectores contiguos la función $F(x_1, x_2, \ldots, x_n)$ cambia de signo.

El teorema demostrado se emplea con frecuencia cuando se elige sólo una fila o bien una columna. El determinante de una matriz de orden 1 coincide con su único elemento. Por ello, el menor dispuesto en la intersección de la i-ésima fila y la j-ésima columna equivale al elemento a_{ij}. Designemos mediante A_{ij} el complemento algebraico del elemento a_{ij}. De acuerdo con el teorema de Laplace, para todo i se tiene

$$a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in} = d. \qquad (40.5)$$

Esta fórmula se llama *desarrollo del determinante por la i-ésima fila*. Análogamente, para todo j

$$a_{1j}A_{1j} + a_{2j}A_{2j} + \ldots + a_{nj}A_{nj} = d, \qquad (40.6)$$

lo que nos proporciona *el desarrollo del determinante por la j-ésima columna*.

En el desarrollo (40.5) sustituyamos los elementos de la i-ésima fila por una totalidad de los n números arbitrarios $b_1, b_2, \ldots, b_n$. La expresión

$$b_1A_{i1} + b_2A_{i2} + \ldots + b_nA_{in}$$

representa el desarrollo del determinante

$$\det \begin{pmatrix} a_{11} a_{12} \ldots a_{1n} \\ \ldots\ldots\ldots \\ b_1 b_2 \ldots b_n \\ \ldots\ldots\ldots \\ a_{n1} a_{n2} \ldots a_{nn} \end{pmatrix} \qquad (40.7)$$

por la i-ésima fila. Este último se obtiene del determinante d sustituyendo la i-ésima fila por una fila de los números $b_1, b_2, \ldots, b_n$. Tomemos ahora, a título de dichos números, los elementos de la k-ésima fila del determinante d para $k \neq i$. El determinante correspondiente (40.7) es igual a cero, ya que tiene dos filas idénticas. Por consiguiente,

$$a_{k1}A_{i1} + a_{k2}A_{i2} + \ldots + a_{kn}A_{in} = 0, \ k \neq i. \qquad (40.8)$$

Por analogía,

$$a_{1k}A_{1j} + a_{2k}A_{2j} + \ldots + a_{nk}A_{nj} = 0, \ k \neq j. \qquad (40.9)$$

Así pues, la suma de los productos de todos los elementos de cualquier fila (columna) de un determinante por los complementos algebraicos de los elementos correspondientes de otra fila (columna) del mismo determinante es igual a cero.

Observemos, como conclusión, que toda la teoría de los determinantes es aplicable, sin cambio alguno, al caso de las matrices complejas. Lo único que se pierde es la claridad relacionada con el concepto de volumen.

Ejercicios.

1. Escríbanse las expresiones de los determinantes de los órdenes segundo y tercero en términos de los elementos de las matrices. Compárense con la expresión (34.6).
2. Escríbase la desigualdad de Hadamard para el determinante de las matrices A y A'.
3. Un determinante de n-ésimo orden, todos los elementos del cual son iguales en módulo a la unidad, equivale a $n^{n/2}$. Demuéstrese que sus filas (columnas) forman una base ortogonal.
4. ¿A qué es igual un determinante, si sus elementos satisfacen las condiciones $a_{ij} = 0$ para $i > j$ ($i < j$, $i \geqslant j$, $i \leqslant j$)?
5. Los elementos de un determinante satisfacen las condiciones $a_{ij} = 0$ para $i > k$ y $j \leqslant k$. Demuéstrese que el determinante es igual al producto del menor principal de orden k por su menor complementario.
6. Supongamos que los elementos de una matriz compleja A satisfacen las condiciones $a_{ij} = \bar{a}_{ji}$ para cualesquiera i, j. Demuéstrese que el determinante de esta matriz es un número real.

§ 41. Dependencia lineal y determinantes

Las más difundidas aplicaciones de los determinantes las encontramos en los problemas relacionados con la dependencia lineal. Supongamos que en un espacio K_n de dimensión n vienen dados m vectores $x_1, x_2, \ldots, x_m$ y es necesario determinar la base de estos vectores. Elijamos una base cualquiera

en K_n y consideremos una tabla rectangular

$$A = \begin{pmatrix} a_{11} a_{12} \dots a_{1n} \\ a_{21} a_{22} \dots a_{2n} \\ \dots\dots\dots \\ a_{m1} a_{m2} \dots a_{mn} \end{pmatrix}. \qquad (41.1)$$

cuyas filas representan las coordenadas de los vectores dados en la base elegida.

Esta tabla se denominará *matriz rectangular*. El primer índice del elemento a_{ij} significa, como hasta ahora, el número de la fila de la matriz en la que se halla dicho elemento, el segundo índice significa el número de la columna. Si se desea subrayar la cantidad de filas y columnas en la matriz A, se escribirá $A\ (m \times n)$ y de la matriz A se dirá que es una matriz $m \times n$. La matriz $A\ (n \times n)$ se llamará, como antes, matriz cuadrada de orden n. Conjuntamente con la matriz A consideraremos la matriz transpuesta A'. Si las dimensiones de A son $m \times n$, las dimensiones de A' serán $n \times m$.

En una matriz rectangular $A\ (m \times n)$ pueden también indicarse diferentes menores cuyo orden no es superior, naturalmente, al menor de los números m, n. Si una matriz tiene no sólo elementos nulos, el orden superior r de los menores, distintos de cero, se denominará *rango* de la matriz A. Cualquier menor de rango r distinto de cero se llama menor *básico*; *básicas* serán también las filas y las columnas en las cuales se encuentra el menor básico. Está claro que pueden haber varios menores básicos. El rango de la matriz nula es igual a cero lo que se desprende de la definición.

Consideraremos que las filas de la matriz A son vectores. Es evidente que si determinamos la base de estos vectores filas, los vectores correspondientes del espacio K_n formarán la base de los vectores $x_1, x_2, \dots, x_m$.

TEOREMA 41.1. *Cualesquiera filas básicas forman la base de los vectores filas de la matriz.*

DEMOSTRACIÓN. Para convencerse de la validez del teorema hace falta mostrar que las filas básicas son linealmente independientes y toda fila de la matriz se expresa linealmente en términos de estas últimas.

Si las filas básicas fueran linealmente dependientes, entonces una de estas filas se expresaría linealmente en términos de las filas básicas restantes. Pero en este caso el menor debe ser igual a cero lo que contradice la hipótesis.

Ahora agreguemos a las filas básicas otra fila cualquiera de la matriz A. Entonces, por definición del menor básico, todos los menores de orden $r + 1$, dispuestos en dichas filas, serán iguales a cero. Supongamos que las filas citadas son linealmente independientes. Al completarlas hasta obtener una base, habremos construido

cierta matriz cuadrada cuyo determinante no debe ser nulo. Pero, por otro lado, desarrollando este determinante por $r+1$ filas iniciales llegamos a la conclusión de que es igual a cero. La contradicción obtenida es testimonio de que toda fila de la matriz A se expresa linealmente en términos de las filas básicas.

El teorema demostrado permite reducir el problema de búsqueda de la base de un sistema de vectores a la búsqueda del menor básico de la matriz. Puesto que el determinante de la matriz transpuesta coincide con el de la matriz inicial, está claro que el teorema 41.1 es válido no sólo para las filas, sino también para las columnas. Esto significa que para cualquier matriz rectangular el rango de su sistema de sus vectores filas es igual al rango de su sistema de vectores columnas. *Lo expuesto no es obvio, si se tiene en cuenta sólo el concepto de rango de un sistema de vectores.*

En un espacio dotado de un producto escalar la dependencia o independencia lineal del sistema de vectores $x_1, x_2, \ldots, x_m$ puede establecerse sin recurrir al desarrollo por la base. Examinemos un determinante

$$G(x_1, x_2, \ldots, x_m) = \det \begin{pmatrix} (x_1, x_1) & (x_1, x_2) & \ldots & (x_1, x_m) \\ (x_2, x_1) & (x_2, x_2) & \ldots & (x_2, x_m) \\ \ldots & \ldots & \ldots & \ldots \\ (x_m, x_1) & (x_m, x_2) & \ldots & (x_m, x_m) \end{pmatrix},$$

que se llama *determinante de Gram* del sistema de vectores $x_1, x_2, \ldots, x_m$.

TEOREMA 41.2. *Un sistema de vectores es linealmente dependiente cuando, y sólo cuando, su determinante de Gram es igual a cero.*

DEMOSTRACIÓN. Supongamos que el sistema de vectores $x_1, x_2, \ldots, x_m$ es linealmente dependiente. En este caso existen tales números $\alpha_1, \alpha_2, \ldots, \alpha_m$, no todos iguales a cero, que

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m = 0.$$

Al multiplicar esta igualdad escalarmente por x_i para todo valor de i, concluimos que las columnas del determinante de Gram son también linealmente dependientes, en otras palabras, el propio determinante es igual a cero.

Supongamos ahora que el determinante de Gram es nulo. Entonces sus columnas son linealmente dependientes, es decir, existen tales números $\alpha_1, \alpha_2, \ldots, \alpha_m$, no todos iguales a cero, que

$$\alpha_1 (x_i, x_1) + \alpha_2 (x_i, x_2) + \ldots + \alpha_m (x_i, x_m) = 0$$

para todo i. Escribamos estas igualdades de la forma siguiente:

$$(x_i, \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m) = 0.$$

Multiplicándolas término a término por $\alpha_1, \alpha_2, \ldots, \alpha_m$ y sumándolas, obtenemos

$$|\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m|^2 = 0.$$

Esto significa que

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m = 0,$$

es decir, que los vectores $x_1, x_2, \ldots, x_m$ son linealmente dependientes.

Ejercicios.

1. ¿Qué representa en sí una matriz, todos los menores de la cual son nulos?
2. ¿Serán las filas básicas y columnas básicas sistemas equivalentes de vectores para una matriz cuadrada?
3. ¿Cambiarán el rango de las matrices las transformaciones elementales examinadas en el § 15?
4. Demuéstrese la desigualdad

$$0 \leqslant G(x_1, x_2, \ldots, x_n) \leqslant \prod_{i=1}^{n} (x_i, x_i).$$

¿Cuáles son los casos en que se consiguen igualdades?

5. Es evidente que

$$\det \begin{pmatrix} \sqrt{2} & 1 \\ 2 & \sqrt{2} \end{pmatrix} = 0.$$

Demuéstrese que con cualquier aproximación del número $\sqrt{2}$ por medio de una fracción decimal p

$$\det \begin{pmatrix} p & 1 \\ 2 & p \end{pmatrix} \neq 0$$

§ 42. Cálculo de los determinantes

El cálculo directo de un determinante, durante el cual se usa la expresión explícita de éste en términos de los elementos de la matriz, es de rara aplicación en la práctica por ser muy laborioso. Un determinante de orden n se compone de $n!$ términos, mientras que para calcular cualquier término y sumarlo con los otros es preciso efectuar n operaciones aritméticas. Si todos estos cálculos se realizaran con ayuda de una computadora moderna, capaz de ejecutar 10^6 operaciones aritméticas por segundo, incluso en este caso, para calcular un determinante de orden 100, por ejemplo, tendríamos que esperar el resultado durante varios millones de años.

Uno de los métodos más efectivos para calcular determinantes está basado en la siguiente idea. Supongamos que en la matriz A existe un elemento a_{kp} distinto de cero. Llamémoslo *elemento rector*. Si a toda i-ésima fila, $i \neq k$, agregamos la k-ésima fila multiplicada por un número arbitrario α_i, el determinante, como se sabe,

no variará. Tomemos

$$\alpha_i = -\frac{a_{ip}}{a_{kp}}$$

y llevemos a cabo el procedimiento citado para todo $i \neq k$. Entonces, en la matriz nueva todos los elementos de la p-ésima columna, a excepción del vector, serán iguales a cero. Desarrollando el determinante nuevo por la p-ésima columna, reduzcamos el cálculo del determinante de n-ésimo orden al cálculo de un solo determinante de orden $(n-1)$. Con este último procedamos de un modo análogo, etc.

El algoritmo descrito se llama *método de Gauss*. Para calcular un determinante de n-ésimo orden, rigiéndose por este método, se requiere cumplir en total alrededor de $\frac{2}{3}n^3$ operaciones aritméticas. De esta manera un determinante de orden 100 puede ser calculado en menos de un segundo en una máquina computadora que realiza 10^6 operaciones aritméticas por segundo.

En conclusión observamos que bajo las condiciones en que las operaciones aritméticas se calculan aproximadamente y los datos se prefijan también de modo aproximado, los resultados de cálculos de los determinantes deben tratarse con cierta cautela. Si las deducciones sobre la dependencia o independencia lineal de un sistema de vectores se basan sólo en el hecho de si es o no igual a cero el determinante, entonces en presencia de la inestabilidad de la que se ha tratado en el § 22, las deducciones pueden resultar erróneas. Al operar con los determinantes, esto siempre debe tenerse en cuenta.

Ejercicios.

1. ¿A qué se debe que los cálculos de un determinante por el método de Gauss son más rápidos que los cálculos directos?

2. Supongamos que todos los elementos de un determinante no sobrepasan, en módulo, la unidad y que al calcular cada término del mismo se comete un error del orden ε. ¿Con qué n el cálculo directo del determinante tendrá sentido desde el punto de vista de su precisión?

3. Constrúyase el algoritmo, basándose en el método de Gauss, para calcular el rango de una matriz rectangular. ¿Qué significa la aplicación de este algoritmo a las condiciones de cálculos aproximados?

CAPÍTULO 5 LÍNEA RECTA Y PLANO EN EL ESPACIO LINEAL

§ 43. Ecuaciones de la línea recta y del plano

El objeto principal de nuestras investigaciones inmediatas serán una recta y un plano en los espacios de segmentos dirigidos. Si fijamos cierto sistema de coordenadas, las coordenadas de los puntos ubicados en una línea recta o en un plano ya no pueden ser arbitrarias, sino que han de satisfacer correlaciones determinadas. Pasamos ahora a la deducción de dichas correlaciones.

Fijemos en un plano el sistema rectangular cartesiano de coordenadas Oxy y una recta L. Consideraremos un vector no nulo

$$n = (A,\ B), \tag{43.1}$$

que es perpendicular a L. Evidentemente, todos los demás vectores perpendiculares a dicha recta serán colineales a n.

Elijamos un punto arbitrario $M_0\,(x_0,\ y_0)$ en la línea recta. Todos los puntos $M\,(x,\ y)$ de la recta L, y sólo ellos, poseen la propiedad de perpendicularidad de los vectores $\overrightarrow{M_0M}$ y n, es decir,

$$(\overrightarrow{M_0M},\ n) = 0. \tag{43.2}$$

Puesto que

$$\overrightarrow{M_0M} = (x - x_0,\ y - y_0),$$

entonces, de (43.1), (43.2) se deduce que

$$A\,(x - x_0) + B\,(y - y_0) = 0.$$

Al introducir la designación

$$-Ax_0 - By_0 = C,$$

concluimos que en el sistema dado Oxy las coordenadas de los puntos de la recta L, y sólo ellas, satisfacen la ecuación

$$Ax + By + C = 0. \tag{43.3}$$

Entre los números A, B hay uno que no es igual a cero. Por esto, llamaremos la ecuación (43.3) ecuación de *primer grado* respecto a las variables x, y.

Demostremos ahora que toda ecuación de primer grado (43.3) define una línea recta con relación al sistema fijado de coordenadas Oxy. Puesto que la ecuación (43.3) es de primer grado, entonces de las constantes A, B aunque una es distinta de cero. Por consiguiente, la ecuación (43.3) tiene al menos una solución x_0, y_0, por ejemplo,

$$x_0 = -\frac{AC}{A^2+B^2}, \qquad y_0 = -\frac{BC}{A^2+B^2},$$

siendo en este caso

$$Ax_0 + By_0 + C = 0.$$

Al sustraer de la ecuación (43.3) la identidad dada, obtendremos la ecuación

$$A(x - x_0) + B(y - y_0) = 0,$$

equivalente a la ecuación (43.3). Mas, esto significa que cualquier punto $M(x, y)$, cuyas coordenadas satisfacen la ecuación (43.3), se halla en una recta que pasa por el punto $M_0(x_0, y_0)$ y es perpendicular al vector (43.1).

Así pues, si tenemos un sistema fijado de coordenadas en un plano, cualquier ecuación de primer grado define una recta y las coordenadas de los puntos de cada recta satisfacen una ecuación de primer grado. La ecuación (43.3) se denomina *ecuación general de la línea recta en un plano*; el vector n de (43.1) se llama *vector normal* de la recta.

Las investigaciones de un plano en un espacio se realizan sin introducir los cambios importantes. Fijemos un sistema rectangular cartesiano de coordenadas $Oxyz$ y consideremos el plano π. Tomemos de nuevo un vector no nulo

$$n = (A, B, C), \tag{43.4}$$

perpendicular a π. Repitiendo los razonamientos anteriores llegamos a la conclusión que todos los puntos $M(x, y, z)$ del plano π, y sólo ellos, satisfacen la ecuación

$$Ax + By + Cz + D = 0, \tag{43.5}$$

que se llamará también ecuación de *primer grado* respecto de las variables x, y, z.

Al examinar nuevamente la ecuación arbitraria de primer grado (43.5), descubriremos que ésta tiene también por lo menos una solución x_0, y_0, z_0, por ejemplo,

$$x_0 = -\frac{AD}{A^2+B^2+C^2}, \qquad y_0 = -\frac{BD}{A^2+B^2+C^2},$$

$$z_0 = -\frac{CD}{A^2+B^2+C^2}.$$

A continuación establecemos que cualquier punto $M(x, y, z)$, cuyas coordenadas satisfacen la ecuación dada (43.5), se dispone en un plano que pasa por el punto $M_0(x_0, y_0, z_0)$ y es perpendicular al vector (43.4).

De este modo, si se fija un sistema de coordenadas en un espacio, toda ecuación de primer grado define un plano y las coordenadas de los puntos de cualquier plano satisfacen la ecuación de primer grado. La ecuación (43.5) se llama *ecuación general del plano en el espacio*; el vector n de (43.4) se denomina vector *normal* del plano.

Veamos ahora cómo se relacionan dos ecuaciones generales que definen una misma línea recta o un plano. Para concretar, sean dadas dos ecuaciones del plano π

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0. \qquad (43.6)$$

Los vectores

$$n_1 = (A_1, B_1, C_1), \qquad n_2 = (A_2, B_2, C_2)$$

son perpendiculares a un mismo plano π y, por ende, son colineales. Como que son, además, no nulos, existe un número t tal que, por ejemplo,

$$n_1 = tn_2$$

o bien

$$A_1 = tA_2, \qquad B_1 = tB_2, \qquad C_1 = tC_2. \qquad (43.7)$$

Multiplicando la segunda ecuación de (43.6) por t y restando de ella la primera, en virtud de las correlaciones (43.7), obtendremos

$$D_1 = tD_2.$$

Por consiguiente, *los coeficientes de las ecuaciones generales que definen una misma recta o un plano son proporcionales.*

Una ecuación general se llama *completa*, si todos sus coeficientes son distintos de cero. Una ecuación no completa se llama *incompleta*. Examinemos la ecuación completa de la recta (43.3). Como todos los coeficientes son distintos de cero, la ecuación puede escribirse en la siguiente forma:

$$\frac{x}{-\frac{C}{A}} + \frac{y}{-\frac{C}{B}} = 1.$$

Al designar

$$a = -\frac{C}{A}, \quad b = -\frac{C}{B},$$

obtendremos una ecuación nueva de la recta

$$\frac{x}{a} + \frac{y}{b} = 1.$$

Esta ecuación lleva el nombre de *ecuación de la recta "en segmentos"*. Los números a, b tienen un sentido geométrico muy simple. Ellos expresan las magnitudes de los segmentos cortados por la recta en los semiejes de coordenadas (fig. 43.1). Por supuesto, la ecuación completa de un plano puede reducirse a una forma análoga

$$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1.$$

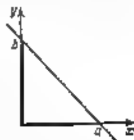


Fig. 43.1.

Las diferentes ecuaciones incompletas definen ciertos casos particulares de la disposición de la recta o del plano. Es útil recordarlas, pues, se encuentran con frecuencia. Por ejemplo, cuando $C = 0$, la ecuación (43.3) define la recta que pasa por el origen de coordenadas; cuando $B = C = 0$, la recta coincide con el eje Oy, etc. Cuando $A = 0$, la ecuación (43.5) define un plano paralelo al eje Ox; cuando $A = B = D = 0$, el plano coincide con el plano coordenado Oxy, etc.

Todo vector no nulo, paralelo a una línea recta, se llamará *vector director*. Consideremos, por ejemplo, el caso de un espacio y hallemos la ecuación de la recta que pasa por un punto dado $M_0(x_0, y_0, z_0)$ y tiene un vector director dado

$$q = (l, m, n).$$

Evidentemente, el punto $M(x, y, z)$ se ubica en la recta citada cuando, y sólo cuando, los vectores $\overrightarrow{M_0M}$ y q son colineales, es decir, cuando, y sólo cuando, las coordenadas de estos vectores son proporcionales, es decir,

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}. \tag{43.8}$$

Estas son las ecuaciones buscadas de la recta. Se denominan comúnmente *ecuaciones canónicas de la recta*. Está claro que en el caso de la recta en un plano la ecuación tendrá la forma:

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}, \tag{43.9}$$

si la recta pasa por el punto $M_0(x_0, y_0)$ y tiene un vector director $q = (l, m)$.

De las ecuaciones canónicas se obtienen con facilidad las ecuaciones de la recta que pasa por dos puntos prefijados M_0, M_1. Con este fin es suficiente tomar, como vector director, el vector $\overrightarrow{M_0M_1}$, expresando sus coordenadas en términos de las coordenadas de los puntos M_0, M_1 y sustituyéndolas en las ecuaciones (43.8), (43.9).

Por ejemplo, en el caso de una recta en el plano tendremos la ecuación siguiente:

$$\frac{x-x_0}{x_1-x_0}=\frac{y-y_0}{y_1-y_0},$$

y en el caso de un espacio:

$$\frac{x-x_0}{x_1-x_0}=\frac{y-y_0}{y_1-y_0}=\frac{z-z_0}{z_1-z_0}.$$

Observemos que en las ecuaciones canónicas de una recta los denominadores pueden resultar ser nulos. Por esto, toda proporción $a/b = c/d$ se entenderá en lo sucesivo como la igualdad $ad = bc$. Por consiguiente, la anulación de una de las coordenadas del vector director significa la reducción a cero del numerador correspondiente en las ecuaciones canónicas.

Con el objeto de representar una recta analíticamente, las coordenadas de sus puntos se escriben, a menudo, como funciones de cierto parámetro auxiliar t. Tomemos por t cada una de las razones iguales en (43.8) y (43.9). Entonces, en el caso de un espacio tendremos las siguientes ecuaciones para una recta:

$$\begin{aligned} x &= x_0 + lt,\\ y &= y_0 + mt,\\ z &= z_0 + nt \end{aligned} \qquad (43.10)$$

y las ecuaciones análogas en el caso de una recta en el plano

$$\begin{aligned} x &= x_0 + lt,\\ y &= y_0 + mt \end{aligned} \qquad (43.11)$$

Las últimas se denominan *ecuaciones paramétricas de una recta*. Asignando al parámetro t diferentes valores, obtendremos diferentes puntos de la recta.

Grandes posibilidades y comodidades en la inscripción de diferentes ecuaciones de una recta y de un plano ofrece el uso del concepto de determinante. Deduzcamos, por ejemplo, la ecuación de un plano que pasa por tres diferentes puntos no dispuestos en una misma recta. Así pues, sean dados tres puntos $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, $M_3(x_3, y_3, z_3)$. Como estos puntos no se hallan en una misma recta, entonces los vectores

$$\overrightarrow{M_1M_2}=(x_2-x_1,\ y_2-y_1,\ z_2-z_1),\ \overrightarrow{M_1M_3}=(x_3-x_1,\ y_3-y_1,\ z_3-z_1)$$

no son colineales. Por esta razón el punto $M(x, y, z)$ se halla en un mismo plano con los puntos M_1, M_2, M_3 cuando, y sólo cuando, los vectores $\overrightarrow{M_1M_2}$, $\overrightarrow{M_1M_3}$ y

$$\overrightarrow{M_1M}=(x-x_1,\ y-y_1,\ z-z_1)$$

son coplanares, es decir, cuando, y sólo cuando, el determinante compuesto de sus coordenadas es igual a cero. Por lo tanto, la ecuación

$$\det \begin{pmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{pmatrix} = 0 \qquad (43.12)$$

es la ecuación del plano buscado que pasa por los tres puntos dados.

Consideremos, por fin, la ecuación de la recta en un espacio la cual es perpendicular a dos rectas no paralelas y pasa por un punto dado. Supongamos que ambas rectas están dadas por sus ecuaciones canónicas

$$\frac{x-x_1}{l_1} = \frac{y-y_1}{m_1} = \frac{z-z_1}{n_1},$$

$$\frac{x-x_2}{l_2} = \frac{y-y_2}{m_2} = \frac{z-z_2}{n_2}.$$

El vector director q de la recta buscada debe ser perpendicular a dos vectores

$$q_1 = (l_1, m_1, n_1), \qquad q_2 = (l_2, m_2, n_2).$$

Estos vectores no son colineales y, por tanto, a título de q puede tomarse, por ejemplo, el producto vectorial $[q_1, q_2]$. Recordando la expresión de las coordenadas de un producto vectorial en términos de las coordenadas de los factores y, usando para su inscripción los determinantes de segundo orden, obtenemos

$$q = \left(\det \begin{pmatrix} m_1 n_1 \\ m_2 n_2 \end{pmatrix}, \quad \det \begin{pmatrix} n_1 l_1 \\ n_2 l_2 \end{pmatrix}, \quad \det \begin{pmatrix} l_1 m_1 \\ l_2 m_2 \end{pmatrix}\right).$$

Si la recta buscada pasa por el punto M_0 (x_0, y_0, z_0), las ecuaciones canónicas de la recta serán:

$$\frac{x-x_0}{\det \begin{pmatrix} m_1 n_1 \\ m_2 n_2 \end{pmatrix}} = \frac{y-y_0}{\det \begin{pmatrix} n_1 l_1 \\ n_2 l_2 \end{pmatrix}} = \frac{z-z_0}{\det \begin{pmatrix} l_1 m_1 \\ l_2 m_2 \end{pmatrix}}.$$

Naturalmente, desde el punto de vista de principio, muchas conclusiones respecto a la ecuación de una recta o de un plano quedan en vigor, cuando se emplea cualquier sistema afín de coordenadas. Nuestro deseo de utilizar los sistemas rectangulares cartesianos de coordenadas se debe, en lo principal, a que los razonamientos, en este último caso, son más simples.

Ejercicios.

1. Escríbase la ecuación de la recta en un plano, que pasa por dos puntos dados, utilizando el determinante de segundo orden. Compárese con (43.12).
2. ¿Será justa la afirmación de que la ecuación (43.12) representa siempre la ecuación de un plano?

3. Escríbanse, por analogía con las ecuaciones (43.10), las ecuaciones paramétricas de un plano en un espacio. ¿Cuántos parámetros deben contener estas ecuaciones?

4. Hállense las coordenadas del vector normal de un plano que pasa por tres puntos dados no dispuestos en una misma recta.

5. ¿Qué representa en el el lugar geométrico de los puntos en un espacio cuyas coordenadas son las soluciones de un sistema de dos ecuaciones lineales algebraicas con tres incógnitas?

§ 44. Disposición conjunta

Cuando se consideran simultáneamente varias rectas y varios planos surgen diferentes problemas relacionados, en primer lugar, con la necesidad de determinar su disposición mutua.

Supongamos que en un espacio dos planos que se cortan están dados por sus ecuaciones generales

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0.$$

Estos planos forman dos ángulos adyacentes que en suma dan dos rectos. Hallemos uno de ellos. Los vectores $n_1 = (A_1, B_1, C_1)$ y $n_2 = (A_2, B_2, C_2)$ son normales y, por ende, la determinación del ángulo entre los planos se reduce a la determinación del ángulo φ entre los vectores n_1, n_2. Conforme a (25.5) tenemos

$$\cos\varphi = \frac{A_1A_2 + B_1B_2 + C_1C_2}{(A_1^2 + B_1^2 + C_1^2)^{1/2}\,(A_2^2 + B_2^2 + C_2^2)^{1/2}}.$$

Por analogía completa se deduce la fórmula para hallar el ángulo entre dos rectas en un plano, dadas por sus ecuaciones generales

$$A_1x + B_1y + C_1 = 0,$$
$$A_2x + B_2y + C_2 = 0.$$

Uno de los ángulos φ, formados por dichas rectas, se calcula según la fórmula

$$\cos\varphi = \frac{A_1A_2 + B_1B_2}{(A_1^2 + B_1^2)^{1/2}\,(A_2^2 + B_2^2)^{1/2}}.$$

La condición de paralelismo de unas rectas, dadas por sus ecuaciones generales, es la condición de carácter colineal de los vectores normales, es decir, la condición de proporcionalidad de sus coordenadas

$$\frac{A_1}{A_2} = \frac{B_1}{B_2}.$$

La condición de perpendicularidad de las rectas coincide con la condición $\cos\varphi = 0$, o bien, lo que es igual, con la condición

$$A_1A_2 + B_1B_2 = 0.$$

Por supuesto, una forma análoga la tiene la condición

$$\frac{A_1}{A_2}=\frac{B_1}{B_2}=\frac{C_1}{C_2}$$

de paralelismo de los planos y la condición

$$A_1A_2 + B_1B_2 + C_1C_2 = 0$$

de perpendicularidad de los planos dados también mediante las ecuaciones generales.

Supongamos ahora que dos rectas en el espacio, por ejemplo, están dadas mediante las ecuaciones canónicas

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1},$$

$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}.$$

Puesto que los vectores $q_1 = (l_1, m_1, n_1)$, $q_2 = (l_2, m_2, n_2)$ son los vectores directores para estas rectas, concluimos de nuevo que uno de los ángulos φ entre las rectas coincidirá con el ángulo entre los vectores q_1, q_2. Por consiguiente,

$$\cos\varphi = \frac{l_1l_2+m_1m_2+n_1n_2}{(l_1^2+m_1^2+n_1^2)^{1/2}\,(l_2^2+m_2^2+n_2^2)^{1/2}}.$$

Correspondientemente, la proporcionalidad de las coordenadas

$$\frac{l_1}{l_2}=\frac{m_1}{m_2}=\frac{n_1}{n_2}$$

es la condición de paralelismo de las rectas y la igualdad

$$l_1l_2 + m_1m_2 + n_1n_2 = 0$$

es la condición de perpendicularidad de las mismas.

Está claro que si las rectas y los planos vienen dados mediante un método en el que se indica de modo explícito el vector director o el normal, entonces la determinación del ángulo entre las rectas y los planos se reduce siempre a la del ángulo entre dichos vectores. Supongamos, por ejemplo, que en un espacio se ha dado el plano π, mediante su ecuación general

$$Ax + By + Cz + D = 0,$$

y la recta L, mediante la ecuación canónica

$$\frac{x-x_0}{l}=\frac{y-y_0}{m}=\frac{z-z_0}{n}.$$

Puesto que el ángulo φ entre la recta y el plano es complementario al ángulo ψ entre el vector director de la recta y el vector nor-

mal del plano (fig. 44.1), entonces

$$\operatorname{sen} \varphi = \frac{|Al + Bm + Cn|}{(A^2 + B^2 + C^2)^{1/2} (l^2 + m^2 + n^2)^{1/2}}.$$

Es evidente la condición

$$Al + Bm + Cn = 0$$

de paralelismo entre la recta y el plano y la condición

$$\frac{A}{l} = \frac{B}{m} = \frac{C}{n}$$

de perpendicularidad de la recta respecto del plano.

La forma de ecuaciones generales, mediante la cual se dan una recta y un plano, permite resolver con toda eficacia un importante problema, a saber, el cálculo de la distancia desde un punto hasta la recta, o bien desde un punto hasta el plano. La deducción de las fórmulas es, en ambos casos, enteramente análoga y nos limitamos de nuevo a la consideración detallada de una sola de ellas.

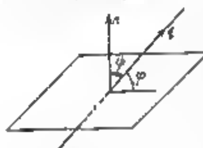


Fig. 44.1.

Supongamos que el plano π en el espacio está dado por su ecuación general (43.5). Tomemos un punto arbitrario $M_0(x_0, y_0, z_0)$. Del punto M_0 tracemos una perpendicular al plano y designemos con $M_1(x_1, y_1, z_1)$ la base de ésta. Está claro que la distancia $\rho(M_0, \pi)$ del punto M_0 al plano es igual a la longitud del vector $\overrightarrow{M_0M_1}$.

Los vectores $n = (A, B, C)$ y $\overrightarrow{M_0M_1} = (x_1 - x_0, y_1 - y_0, z_1 - z_0)$ son perpendiculares al mismo plano, razón por la cual son colineales. Por eso, existe un número t tal que $\overrightarrow{M_0M_1} = tn$, es decir,

$$x_1 - x_0 = tA,$$
$$y_1 - y_0 = tB,$$
$$z_1 - z_0 = tC.$$

El punto $M_1(x_1, y_1, z_1)$ se halla en el plano π. Al expresar sus coordenadas a base de las correlaciones obtenidas y sustituirlas en la ecuación del plano, tenemos

$$t = -\frac{Ax_0 + By_0 + Cz_0 + D}{A^2 + B^2 + C^2}.$$

Pero, la longitud del vector n es igual a $(A^2+B^2+C^2)^{1/2}$ y, por ende, $|\overrightarrow{M_0M_1}| = |t|\,(A^2+B^2+C^2)^{1/2}$. Por consiguiente

$$\rho(M_0,\ \pi)=\frac{|Ax_0+By_0+Cz_0+D|}{(A^2+B^2+C^2)^{1/2}}.$$

En particular,

$$\rho(0,\ \pi)=\frac{|D|}{(A^2+B^2+C^2)^{1/2}}.$$

A la par con la ecuación general (43.5) del plano consideraremos, además, las siguientes ecuaciones suyas

$$\pm(A^2+B^2+C^2)^{-1/2}(Ax+By+Cz+D)=0.$$

De los dos signos posibles en el primer miembro elijamos el opuesto al signo de D. Si es que $D=0$, elijamos un signo cualquiera. Entonces, el término independiente de esta ecuación será un número no positivo $-p$, mientras que los coeficientes de x, y, z serán cosenos de los ángulos entre el vector normal y los ejes de coordenadas. La ecuación

$$x\cos\alpha+y\cos\beta+z\cos\gamma-p=0 \qquad (44.1)$$

se llama *ecuación normalizada del plano*. Es evidente que

$$\rho(M_0,\ \pi)=|x_0\cos\alpha+y_0\cos\beta+z_0\cos\gamma-p|,$$
$$\rho(0,\ \pi)=p.$$

La distancia $\rho(M_0,\ L)$ desde el punto $M_0(x_0,\ y_0)$ hasta la recta L en el plano, dada mediante su ecuación general (43.3), se determina por la fórmula análoga

$$\rho(M_0,\ L)=\frac{|Ax_0+By_0+C|}{(A^2+B^2)^{1/2}}.$$

La ecuación normalizada de una línea recta tiene por expresión

$$x\cos\alpha+y\,\mathrm{sen}\,\alpha-p=0. \qquad (44.2)$$

Aquí, α es el ángulo formado por el vector normal con el eje Ox.

Ejercicios.

1. ¿En qué condiciones, impuestas en las coordenadas de los vectores normales, dos rectas en un plano (tres planos en un espacio) se intersecan en un punto?
2. ¿Bajo qué condición la recta (43.8) pertenece al plano (43.5)?
3. ¿Bajo qué condición dos rectas en un espacio, dadas mediante ecuaciones canónicas, pertenecen a un plano?
4. Calcúlense los ángulos entre la diagonal de un cubo y las caras de éste.
5. Dedúzcase la fórmula para calcular la distancia entre un punto y una recta en un espacio, dada esta última mediante sus ecuaciones canónicas.

§ 45. Plano en el espacio lineal

Ya hemos subrayado más de una vez que una recta y un plano que pasan por el origen de coordenadas se identifican en los espacios de los segmentos dirigidos con la imagen geométrica de un subespacio. Pero según sus propiedades, difieren poco de cualesquiera otras rectas y otros planos que se obtienen por traslación paralela o por desplazamiento de estos subespacios. Deseando generalizar este hecho para cualesquiera espacios lineales, llegamos al concepto de plano en el espacio lineal.

Sea L cierto subespacio del espacio lineal K. Fijemos en K un vector arbitrario x_0. En particular, este vector puede pertenecer a L. El conjunto H de vectores z que se obtienen según la fórmula

$$z = x_0 + y, \qquad (45.1)$$

donde y es cualquier vector de L, se denomina *plano* en el espacio lineal K. El vector x_0 se llama *vector de desplazamiento* y el subespacio L, *director*. En cuanto al plano H, diremos que está formado por el desplazamiento del subespacio L al vector x_0.

El concepto formal de un plano incluye en sí las nociones de rectas y planos (en la interpretación vectorial) en los espacios de segmentos dirigidos. Pero, por ahora no sabemos si posee las propiedades análogas.

Todo vector del plano H se representa del modo *único* en forma de la suma (45.1). Si $z = x_0 + y$ y $z = x_0 + y'$, donde $y, y' \in L$, resulta que $y = y'$. De la fórmula (45.1) se deduce, además, que la diferencia entre dos vectores cualesquiera del plano H pertenece al subespacio L.

Elijamos en el plano H un vector arbitrario z_0. Sea $z_0 = x_0 + y_0$. Representamos la igualdad (45.1) en la forma:

$$z = z_0 + (y - y_0).$$

Los conjuntos de vectores y e $y - y_0$ describen el mismo subespacio L. La última igualdad significa, por ello, que el plano H puede obtenerse por desplazamiento del subespacio L a *todo* vector fijado del mismo plano.

El plano H es un conjunto de vectores de K generado por el subespacio L y el vector de desplazamiento x_0, de conformidad con (45.1). Una circunstancia de mucha importancia consiste en que cualquier plano puede ser generado por un solo subespacio. Supongamos que no es así, es decir, que existen un subespacio director más L', y un vector de desplazamiento x_0', que forman el mismo plano H. En tal caso para todo $z \in H$ tenemos $z = x_0 + y$, donde $y \in L$, y al mismo tiempo $z = x_0' + y'$, donde $y' \in L'$. De aquí se desprende que el subespacio L' es un conjunto de vectores de K,

definidos mediante la fórmula

$$y' = (x_0 - x_0') + y.$$

Como el vector nulo pertenece a L', de la última fórmula se infiere que el vector $(x_0 - x_0')$ pertenece a L. Mas, esto significa que el subespacio L' se compone de los mismos vectores que el subespacio L.

Ya hemos indicado que el vector de desplazamiento se define mediante un plano y, además de un modo *no unívoco*. No obstante, en este caso también la cuestión acerca de la univocidad puede resolverse de una manera bien natural.

Convengamos en considerar que en el espacio lineal K se ha introducido un producto escalar. Si en lugar del vector x_0 tomamos $\mathrm{ort}_L x_0$, está claro que obtendremos el mismo plano. Por ello, sin restringir la generalidad, se puede considerar que $x_0 \perp L$. El vector x_0 en este caso se llamará vector *ortogonal* de desplazamiento. Ahora podremos demostrar que todo plano es generado sólo por un vector de desplazamiento.

Efectivamente, supongamos que existen dos vectores de desplazamiento x_0', x_0'' que son ortogonales al subespacio L y generan, sin embargo, un mismo plano H. Entonces, para cualquier vector $y' \in L$ debe existir un vector $y'' \in L$ tal que $x_0' + y' = x_0'' + y''$. De aquí se deduce que $x_0' - x_0'' \in L$. Pero, según la hipótesis, $x_0' - x_0'' \perp L$. Por consiguiente, $x_0' - x_0'' = 0$, es decir, $x_0' = x_0''$.

Esto significa, en particular, que en un espacio dotado de un producto escalar cualquier plano cuenta sólo con un vector ortogonal al subespacio director.

Dos planos se denominan *paralelos*, si el subespacio director de uno de ellos forma parte del subespacio director del otro.

Esta afirmación se justifica con facilidad. Cualesquiera dos planos paralelos H_1, H_2 *o bien no contienen ningún vector común* o bien uno de ellos forma parte del otro. Supongamos que H_1, H_2 tienen un vector común z_0. Puesto que todo plano puede ser obtenido por desplazamiento del subespacio director a cualquier vector suyo, entonces tanto H_1 como H_2 se obtienen por desplazamiento de los subespacios correspondientes al vector z_0. Pero uno de los subespacios forma parte del otro, por lo cual uno de los planos forma parte del otro.

Un subespacio es un caso particular del plano. Es evidente que el subespacio L es paralelo a todo plano H, obtenido por desplazamiento de L a cierto vector x_0. De la propiedad demostrada para los planos paralelos se infiere que H coincide con L cuando, y sólo cuando, $x_0 \in L$.

Consideraremos ahora dos planos no paralelos H_1, H_2. Estos o bien no tienen ningún vector común o bien tienen el vector común.

En el primer caso los planos H_1, H_2 se llamarán planos *que se cruzan*, en el segundo caso llamémoslos planos *que se intersecan*.

Al igual que en el caso de subespacios, un conjunto de vectores pertenecientes simultáneamente a H_1 y H_2 se denominará *intersección* de estos planos y se indicará $H_1 \cap H_2$. Supongamos que el plano H_1 se ha formado por desplazamiento del subespacio L_1 y el plano H_2, por desplazamiento del subespacio L_2. Designaremos

$$H = H_1 \cap H_2, \qquad L = L_1 \cap L_2.$$

TEOREMA 20.1. *Si la intersección H contiene el vector z_0, representa en sí un plano formado por desplazamiento de la intersección L a dicho vector.*

DEMOSTRACION. Según la hipótesis del teorema, existe un vector z_0 perteneciente a la intersección H. Supongamos que existe un vector más $z_1 \in H$. Representémoslo en la forma:

$$z_1 = z_0 + (z_1 - z_0).$$

Ahora, de la ecuación de correlaciones

$$z_1, z_0 \in H \to z_1, z_0 \in H_1; \quad z_1, z_0 \in H_2 \to z_1 - z_0 \in L_2;$$
$$z_1 - z_0 \in L_1 \to z_1 - z_0 \in L$$

concluimos que todo vector de la intersección H puede ser representado como la suma del vector z_0 y cierto vector de la intersección L.

Tomemos, luego, un vector arbitrario f de L. Se tiene

$$f \in L \to f \in L_1; \quad f \in L_2 \to z_0 + f \in H_1;$$
$$z_0 + f \in H_2 \to z_0 + f \in H,$$

es decir, todo vector del subespacio L desplazado al vector z_0, pertenece a la intersección H. El teorema queda demostrado.

Un plano no ha de ser forzosamente un subespacio. No obstante, se le puede atribuir *una dimensión* igual a la del subespacio director. Un plano de dimensión nula contiene sólo un vector: el vector de desplazamiento. Al determinar la dimensión de la intersección de los planos será útil el teorema 19.1. De los teoremas 19.1, 45.1 se desprende que la dimensión de la intersección H no es superior a la mínima de las dimensiones H_1, H_2.

Si en un espacio de segmentos dirigidos vienen dados dos (tres) vectores, entonces, al imponer ciertas condiciones adicionales, podemos construir sólo un plano de dimensión 1 (2) que contenga los vectores dados. Dichas condiciones adicionales pueden enunciarse del modo siguiente. Los dos vectores dados no deben ser coincidentes, es decir, no deben pertenecer a un plano de dimensión nula. Los tres vectores dados no deben pertenecer a un plano de dimensión uno.

Hechos análogos tienen lugar en un espacio lineal arbitrario.

Supongamos que en un espacio lineal están dados los vectores $x_0, x_1, \dots, x_k$. Diremos que estos vectores se encuentran en *la posición general*, si no pertenecen a un plano de dimensión $k-1$.

TEOREMA 45.2. *Si los vectores $x_0, x_1, \dots, x_k$ se encuentran en la posición general, existe un único plano H de dimensión k que los contiene.*

DEMOSTRACION. Examinemos los vectores $x_1 - x_0, x_2 - x_0, \dots, x_k - x_0$. Si fueran linealmente dependientes, pertenecerían a cierto subespacio de dimensión no superior a $k-1$. Por consiguiente, los propios vectores $x_0, x_1, \dots, x_k$ pertenecerían a un plano obtenido por desplazamiento de este subespacio al vector x_0, lo que contradice la hipótesis del teorema.

Así pues, los vectores $x_1 - x_0, x_2 - x_0, \dots, x_k - x_0$ son linealmente independientes. Indiquemos con L su cápsula lineal. El subespacio L tiene una dimensión k. Al desplazarlo al vector x_0, obtendremos cierto plano H de la misma dimensión a la que pertenecen todos los vectores dados $x_0, x_1, \dots, x_k$.

El plano construido H es único. En efecto, supongamos que los vectores $x_0, x_1, \dots, x_k$ pertenecen a dos planos H_1, H_2, ambos de dimensión k. El plano no cambia, si el vector de desplazamiento se sustituye por cualquier otro vector del plano. Por ello, sin restringir la generalidad, se puede considerar que H_1, H_2 se han obtenido por desplazamiento respectivo de los subespacios L_1, L_2 a un mismo vector x_0. Pero en este caso resulta que ambos subespacios coinciden, puesto que tienen la misma dimensión k y contienen un mismo sistema linealmente independiente $x_1 - x_0, x_2 - x_0, \dots, x_k - x_0$.

Ejercicios.

1. Sean H_1, H_2 dos planos cualesquiera. Llamemos *suma* $H_1 + H_2$ *de los planos* H_1, H_2 el conjunto de todos los vectores del tipo $x_1 + x_2$, donde $x_1 \in H_1$, $x_2 \in H_2$. Demuéstrese que la suma de los planos es un plano.

2. Sea H un plano, y sea λ un número. Llamemos *producto* λH *del plano* H *por el número* λ al conjunto de todos los vectores del tipo λx, donde $x \in H$. Demuéstrese que este producto es un plano.

3. ¿Será espacio lineal el conjunto de todos los planos de un mismo espacio con las operaciones sobre ellos introducidas arriba?

4. Demuéstrese que los vectores $x_0, x_1, \dots, x_k$ se encuentran en la posición general cuando, y sólo cuando, los vectores $x_1 - x_0, x_2 - x_0, \dots, x_k - x_0$ son linealmente independientes.

5. Demuéstrese que un plano de dimensión k que contiene los vectores de posición general $x_0, x_1, \dots, x_k$ es un subespacio cuando, y sólo cuando, estos vectores son linealmente dependientes.

§ 46. Recta e hiperplano

En el espacio lineal K de dimensión n existen dos clases de planos que ocupan una posición especial. Se trata de los planos de dimensión 1 y de los planos de dimensión

$m-1$. De acuerdo con la imagen geométrica en los espacios de segmentos dirigidos, todo plano de dimensión 1 se llama *línea recta*. Un plano de dimensión $m-1$ se llama *hiperplano*.

Consideremos una recta arbitraria H en el espacio lineal K. Designemos mediante x_0 el vector de desplazamiento y mediante q, el vector básico del subespacio director unidimensional. Supongamos que estos vectores están dados por sus coordenadas

$$x_0 = (x_1, x_2, \ldots, x_m),$$
$$q = (q_1, q_2, \ldots, q_m)$$

respecto de cierta base del espacio K. Evidentemente, todo vector z de la recta H puede ser dado en la forma:

$$z = x_0 + tq, \qquad (46.1)$$

donde t es un número. Por esto la correlación (46.1) se puede considerar ecuación vectorial de la recta H en el espacio K. Si el vector z tiene en la misma base las coordenadas

$$z = (z_1, z_2, \ldots, z_m),$$

entonces, escribiendo la igualdad (46.1) según las coordenadas, obtendremos

$$\begin{aligned} z_1 &= x_1 + q_1 t, \\ z_2 &= x_2 + q_2 t, \\ &\ldots\ldots\ldots \\ z_m &= x_m + q_m t. \end{aligned} \qquad (46.2)$$

Al comparar ahora estas ecuaciones con (43.10), (43.11), resulta natural denominarlas *ecuaciones paramétricas* de la recta H. Diremos que la recta H pasa por el vector x_0 y tiene el *vector director* q.

De acuerdo con el teorema 45.2, por cualesquiera dos vectores no coincidentes x_0, y_0 siempre se puede trazar una recta y, además, una sola. Supongamos que en cierta base del espacio K los vectores x_0, y_0 vienen dados por sus coordenadas

$$x_0 = (x_1, x_2, \ldots, x_m),$$
$$y_0 = (y_1, y_2, \ldots, y_m).$$

Como que a título de vector director puede tomarse, por ejemplo, el vector $y_0 - x_0$, entonces, de la ecuación (46.2) obtenemos las ecuaciones paramétricas

$$\begin{aligned} z_1 &= x_1 + (y_1 - x_1)\, t, \\ z_2 &= x_2 + (y_2 - x_2)\, t, \\ &\ldots\ldots\ldots\ldots \\ z_m &= x_m + (y_m - x_m)\, t \end{aligned} \qquad (46.3)$$

de una recta que pasa por los dos vectores dados.

Cuando $t = 0$, estas ecuaciones definen el vector x_0; cuando $t = 1$, el vector y_0. Si el espacio K es real, el conjunto de vectores dados mediante las ecuaciones (46.3) para $0 \leqslant t \leqslant 1$ se llamará *segmento* que une los vectores x_0, y_0. Por supuesto, esta denominación está ligada con la imagen geométrica del conjunto dado en los espacios de segmentos dirigidos.

Supongamos que la recta H se corta con cierto plano. En tal caso, de acuerdo con el corolario del teorema 45.1, la intersección será o bien una recta o bien un vector. Si la intersección resulta ser una recta, ésta coincidirá, por supuesto, con la recta H. Pero esto significa que al cortarse una línea recta con un plano, la recta o bien se manifiesta íntegramente en el plano o bien tiene con éste sólo un vector común.

La noción de hiperplano tiene sentido en todo espacio lineal, sin embargo la emplearemos sólo en los espacios dotados de producto escalar.

Examinemos un hiperplano arbitrario H. Supongamos que se ha formado por desplazamiento del subespacio $(m-1)$-dimensional L el vector x_0. El complemento ortogonal $L^{\perp}$ será, en este caso, un subespacio unidimensional. Indiquemos con n cualquiera de sus vectores básicos. El vector z pertenece al hiperplano H cuando, y sólo cuando, el vector $z - x_0$ pertenezca al subespacio L. A su vez, esta condición se cumple cuando, y sólo cuando, el vector $z - x_0$ sea ortogonal al vector n, es decir,

$$(n, z - x_0) = 0. \tag{46.4}$$

De esta modo, hemos obtenido una ecuación la cual se satisface por todos los vectores del hiperplano H. Con el fin de definir un hiperplano en forma de tal ecuación, basta indicar cualquier vector n, ortogonal al subespacio director, y el vector de desplazamiento x_0.

El hecho de que la ecuación tenga una forma explícita permite simplificar, en grado considerable, diversas investigaciones. Sean dados los vectores $n_1, n_2, \ldots, n_k$ y $x_1, x_2, \ldots, x_k$. Investiguemos un plano R que es la intersección de los hiperplanos

$$\begin{aligned} (n_1, z - x_1) &= 0, \\ (n_2, z - x_2) &= 0, \\ &\ldots\ldots\ldots \\ (n_k, z - x_k) &= 0. \end{aligned} \tag{46.5}$$

Este problema puede considerarse como resolución del sistema de ecuaciones (46.5) respecto de los vectores z. Supongamos que la intersección de los hiperplanos no es vacía, es decir, que el sistema (46.5) tiene por lo menos una solución z_0. Entonces, como se sabe,

el plano buscado se determina también por el siguiente sistema:

$$\begin{aligned} &(n_1, z - z_0) = 0,\\ &(n_2, z - z_0) = 0,\\ &\dots\dots\dots\dots\\ &(n_k, z - z_0) = 0 \end{aligned} \tag{46.6}$$

en virtud de que ningún plano varíe, si el vector de desplazamiento se sustituye por cualquier otro vector del plano.

El vector $y = z - z_0$ es un vector arbitrario de la intersección L de los subespacios directores de todos los k hiperplanos. Es evidente que los vectores y del subespacio L satisfacen el siguiente sistema:

$$\begin{aligned} &(n_1, y) = 0,\\ &(n_2, y) = 0,\\ &\dots\dots\dots\\ &(n_k, y) = 0. \end{aligned} \tag{46.7}$$

La intersección L, dada en forma del sistema (46.7), permite resolver con facilidad la cuestión sobre la dimensión de L. Según se desprende del mismo sistema, el subespacio L es un complemento ortogonal de la cápsula lineal del sistema de vectores $n_1, n_2, \dots, n_k$. Si r es el rango de dicho sistema, la dimensión de L y, por lo tanto, de R será igual a $m - r$, donde m es la dimensión del espacio. En particular, si los vectores $n_1, n_2, \dots, n_k$ son linealmente independientes, la dimensión del plano (46.5) es $m - k$. En este caso se presupone, desde luego, que el plano existe, es decir, el sistema (46.5) tiene al menos una solución. Con el fin de prefijar el subespacio L, definido por el sistema (46.7), basta indicar su base, es decir, cualquier sistema de $m - r$ vectores linealmente independientes, ortogonales a los vectores $n_1, n_2, \dots, n_k$.

La ecuación (46.4) del hiperplano puede escribirse también de una forma algo diferente. Designemos $(n, x_0) = b$, entonces la ecuación

$$(n, x) = b \tag{46.8}$$

definirá el mismo hiperplano que la ecuación (46.4). Hemos de notar que las ecuaciones generales (43.3), (43.5) de una recta y de un plano son, en esencia, las mismas ecuaciones. Resulta importante subrayar que *toda* ecuación del tipo (46.8) puede ser reducida al tipo (46.4), eligiendo de una manera adecuada el vector x_0. Para ello es suficiente, por ejemplo, tomar x_0 en la forma:

$$x_0 = \alpha n.$$

Al sustituir esta expresión en (46.4) y comparando con (46.8), concluimos que debe verificarse

$$\alpha = \frac{b}{(n, n)}$$

Ahora podemos hacer una deducción de que si un sistema del tipo (46.5) define cierto plano, el mismo plano puede ser definido también por el sistema del tipo siguiente:

$$\begin{aligned} (n_1, z) &= b_1, \\ (n_2, z) &= b_2, \\ &\ldots\ldots \\ (n_k, z) &= b_k \end{aligned} \qquad (46.9)$$

con los correspondientes números $b_1, b_2, \ldots, b_k$. Evidentemente, será lícita también la afirmación inversa. El sistema (46.9) define el mismo plano que el sistema (46.5) con los vectores correspondientes $z_1, z_2, \ldots, z_k$.

La recta y el plano son hiperplanos en los espacios V_2 y V_3, respectivamente. Anteriormente se ha establecido la relación que existe entre la distancia desde un punto hasta dichos hiperplanos y el resultado de la sustitución de las coordenadas del punto en las ecuaciones generales. Una relación análoga tiene lugar también en el caso de un hiperplano arbitrario.

Sea H un hiperplano dado mediante la ecuación (46.4). Como hasta ahora, designamos con $\rho(v, H)$ la distancia entre el vector v y H. Tomando en consideración la ecuación (46.4), obtenemos que

$$(n, v - z_0) = (n, v - z_0) - (n, z - z_0) = (n, v - z)$$

para cualquier vector z de H. De conformidad con la desigualdad de Cauchy—Buniakovski

$$|(n, v - z_0)| \leqslant |n|\,|v - z|. \qquad (46.10)$$

Por esto

$$|v - z| \geqslant \frac{|(n, v - z_0)|}{|n|}.$$

Si mostramos que en el hiperplano H existe un vector z^* y, además, sólo uno, para el cual la desigualdad (46.10) se convierte en una igualdad, esto será indicio de que, en primer lugar,

$$\rho(v, H) = \frac{|(n, v - z_0)|}{|n|} \qquad (46.11)$$

y, en segundo lugar, el valor $\rho(v, H)$ se logra en un solo vector z^*.

Designemos con L el subespacio director del hiperplano H. Está claro que todo vector, ortogonal a L, será colineal con n, y viceversa. La desigualdad (46.10) se convierte en igualdad cuando, y sólo cuando, los vectores n y $v - z$ son colineales, es decir, $v - z = \alpha n$ para cierto número α. Supongamos que la igualdad se verifica para dos vectores z_1, z_2 de H, es decir,

$$v - z_1 = \alpha_1 n,$$
$$v - z_2 = \alpha_2 n.$$

De aquí se desprende que

$$z_1 - z_2 = (\alpha_2 - \alpha_1)\, n.$$

Por consiguiente, $z_1 - z_2 \perp L$. Pero $z_1 - z_2 \in L$ como la diferencia entre dos vectores del hiperplano. Por ello $z_1 - z_2 = 0$ ó $z_1 = z_2$.

Designemos mediante z_0 un vector de H ortogonal a L. Como se sabe, este vector existe y es único. Escribamos el vector z en forma de la suma

$$z = z_0 + y,$$

donde $y \in L$. Representemos el vector v en la forma

$$v = f + s,$$

donde $f \in L$, $s \perp L$. Ahora

$$v - z = (s - z_0) + (f - y).$$

Si hacemos $z = z_0 + f$,

$$v - z = s - z_0.$$

El vector $h = s - z_0$ es ortogonal a L y la fórmula (46.11) queda establecida.

Al mismo tiempo hemos demostrado que todo vector v del espacio puede ser representado de modo *único* en forma de la suma

$$v = z + h,$$

donde el vector z pertenece al hiperplano H y el vector h es ortogonal al subespacio director L. Por analogía con los espacios de segmentos dirigidos, el vector z en esta descomposición se denomina *proyección* del vector v sobre el hiperplano H; h es *la perpendicular* trazada del vector v sobre H. El proceso en el que el vector z se obtiene de v se llama *proyección* de v sobre H. Si el hiperplano está dado por la ecuación (46.4), el vector n se llama vector *normal* del hiperplano. Para los vectores dados z_0 y n existe *un único* hiperplano que contiene el vector z_0 y es ortogonal al vector n.

Ejercicios.

1. Demuéstrese que cualquier plano distinto de todo el espacio puede ser dado como una intersección de los hiperplanos (46.9).
2. Demuéstrese que la suma de hiperplanos será un hiperplano cuando, y sólo cuando, los hiperplanos sumados son paralelos.
3. Demuéstrese que el producto de un hiperplano por un número no nulo es un hiperplano.
4. ¿Cuáles son las condiciones de paralelismo entre una recta y un hiperplano?
5. Dedúzcase la fórmula de la distancia entre un vector y una recta dada mediante la ecuación (46.1).

§ 47. Semiespacio

Con las nociones de recta e hiperplano de un espacio está relacionada la noción de los así llamados conjuntos convexos. Como que estos conjuntos son de amplio uso en las más diversas ramas de la matemática, fijemos nuestra atención en la investigación de una parte de ellos.

Un conjunto de vectores de un espacio lineal real se denomina *convexo*, si junto con cada dos vectores contiene también todo el segmento que los une.

Como conjuntos convexos pueden intervenir, por ejemplo, un vector, un segmento, una recta, un subespacio, un plano, un hiperplano y varios más.

Supongamos que el hiperplano en un espacio real viene dado mediante la ecuación

$$(n, z) - b = 0.$$

Un conjunto de vectores z que satisfacen la desigualdad

$$(n, z) - b < 0 \tag{47.1}$$

ó

$$(n, z) - b > 0. \tag{47.2}$$

se llama *semiespacio abierto*. El semiespacio (47.1) se denomina *negativo* y el (47.2), *positivo*.

TEOREMA 47.1. *Un semiespacio es un conjunto convexo.*

DEMOSTRACIÓN. Tomemos dos vectores x_0, y_0 y designemos

$$\Phi_1 = (n, x_0) - b, \qquad \Phi_2 = (n, y_0) - b.$$

Si z es un vector cualquiera de una recta que pasa por x_0, y_0, entonces

$$z = x_0 + t(y_0 - x_0).$$

Para $0 \leqslant t \leqslant 1$ obtenemos los vectores del segmento que une x_0, y_0. Tenemos

$$(n, z) - b = \Phi_1 (1 - t) + \Phi_2 t. \tag{47.3}$$

Si Φ_1 y Φ_2 tienen signos idénticos, es decir, si los vectores x_0, y_0 pertenecen a un mismo semiespacio, el segundo miembro de la correlación (47.3) será del mismo signo que Φ_1 y Φ_2 para todos los valores de t que satisfacen las desigualdades $0 \leqslant t \leqslant 1$.

De este modo, todo hiperplano divide el espacio lineal en tres conjuntos convexos disjuntos, a saber, el propio hiperplano y dos semiespacios abiertos.

Supongamos que los vectores x_0, y_0 pertenecen a diferentes semiespacios, es decir, Φ_1 y Φ_2 tienen signos opuestos. La transformación formal de la correlación (47.3) conduce a la desigualdad

siguiente:

$$(n, z) - b = (\Phi_2 - \Phi_1)\left(t - \frac{1}{1 - \Phi_2/\Phi_1}\right).$$

De aquí se infiere que una recta que pasa por los vectores x_0, y_0 corta el hiperplano. La intersección se determina por el valor

$$t = \frac{1}{1 - \Phi_2/\Phi_1},$$

que satisface las desigualdades $0 \leqslant t \leqslant 1$. Así pues,

Si dos vectores pertenecen a diferentes semiespacios el segmento que une dichos vectores corta el hiperplano que define los semiespacios.

Teniendo presente la propiedad enunciada, es fácil comprender qué representan en sí los semiespacios en los espacios de segmentos dirigidos. En un plano los extremos de los vectores del semiespacio se disponen de un lado de la línea recta; en un espacio, de un lado del plano.

A la par con los semiespacios abiertos se consideran, con frecuencia, *los cerrados*. Se definen como conjuntos de los vectores z que satisfacen la desigualdad

$$(n, z) - b \leqslant 0 \qquad (47.4)$$

ó

$$(n, z) - b \geqslant 0. \qquad (47.5)$$

El semiespacio (47.4) se llama *no positivo*, el (47.5) *no negativo*. Por supuesto, los semiespacios cerrados son también conjuntos convexos.

TEOREMA 47.2. *Una intersección de conjuntos convexos es un conjunto convexo.*

DEMOSTRACIÓN. Es suficiente, evidentemente, examinar el caso de dos conjuntos U_1, U_2. Sea $U = U_1 \cap U_2$ su intersección. Tomemos cualesquiera dos vectores x_0, y_0 de U y designemos mediante S el segmento que los une. Los vectores x_0, y_0 pertenecen tanto a U_1 como a U_2. Por ello, por ser los conjuntos U_1, U_2 convexos, el segmento S pertenece íntegramente tanto a U_1 como a U_2, es decir, el segmento S pertenece a la intersección U.

El teorema demostrado es de gran importancia, cuando se estudian los conjuntos convexos. En particular, permite afirmar que un conjunto no vacío de vectores z que satisfacen simultáneamente el sistema de desigualdades

$$\begin{gathered}(n_1, z) - f_1 \gtreqless 0,\\ (n_2, z) - f_2 \gtreqless 0,\\ \dots\dots\dots\dots\\ (n_m, z) - f_m \gtreqless 0\end{gathered}$$

es convexo. Semejantes sistemas de desigualdades sirven de elemento principal en la descripción de muchos problemas referentes a la planificación de la producción, al mando, etc.

Ejercicios.

1. Demuéstrese que el conjunto de vectores z que satisfacen la condición $(n, z) \leqslant \alpha$, es convexo.

2. Demuéstrese que si un vector z pertenece al hiperplano (46.8), el vector $z + n$ se ubica en un semiespacio positivo.

3. Demuéstrese que si los hiperplanos están dados mediante las ecuaciones normales (44.1), (44.2), el origen de coordenadas se dispone siempre en un semiespacio no positivo.

§ 48. Sistemas de ecuaciones lineales

Volvemos nuevamente a la investigación de los sistemas de ecuaciones algebraicas lineales, esta vez en relación con el problema de la intersección de hiperplanos.

Examinemos un espacio K, real o complejo, de dimensión m. Supongamos que en el espacio se ha introducido un producto escalar. Elijamos una base ortonormalizada y admitamos que los vectores normales $n_1, n_2, \ldots, n_k$ de los hiperplanos $H_1, H_2, \ldots, H_k$ de (46.9) están dados mediante sus coordenadas

$$
\begin{aligned}
n_1 &= (a_{11}, a_{12}, \ldots, a_{1m}),\\
n_2 &= (a_{21}, a_{22}, \ldots, a_{2m}),\\
&\ldots\ldots\ldots\ldots\\
n_k &= (a_{k1}, a_{k2}, \ldots, a_{km})
\end{aligned}
\qquad (48.1)
$$

respecto de dicha base. Convengamos en considerar que los vectores

$$z = (z_1, z_2, \ldots, z_m),$$

que pertenecen a la intersección de los hiperplanos, se determinan también por sus coordenadas en la misma base.

En el caso de un espacio real, el producto escalar de vectores en la base ortonormalizada es igual a la suma de los productos de coordenadas tomados dos a dos. Por ello, el sistema (46.9), expresado en coordenadas, tendrá la forma siguiente

$$
\begin{aligned}
a_{11}z_1 + a_{12}z_2 + \ldots + a_{1m}z_m &= b_1,\\
a_{21}z_1 + a_{22}z_2 + \ldots + a_{2m}z_m &= b_2,\\
&\ldots\ldots\ldots\ldots\\
a_{k1}z_1 + a_{k2}z_2 + \ldots + a_{km}z_m &= b_k.
\end{aligned}
\qquad (48.2)
$$

En el caso de un espacio complejo obtendremos de nuevo un sistema análogo, a excepción de que los coeficientes entre las incógnitas y los

segundos *miembros* se sustituirán por unos números conjugados complejos.

Así pues, el problema de la intersección de hiperplanos se reduce a otro problema en el que se resuelve un sistema de ecuaciones algebraicas lineales y que ya estudiamos en el párrafo 22. Evidentemente, cualquier sistema de ecuaciones de coeficientes complejos o reales puede también investigarse desde el punto de vista de la intersección de unos hiperplanos en el espacio complejo o real P_m.

Una cuestión de importancia es la investigación de un sistema de ecuaciones algebraicas lineales relacionada con *la compatibilidad* de éste. Precisamente con esta cuestión está ligada la respuesta a la pregunta sobre si es conjunto vacío o no la intersección de hiperplanos. Desde luego, para realizar tal investigación se puede aprovechar el método de Gauss. No obstante, este método no es siempre cómodo.

Al estudiar los sistemas de ecuaciones algebraicas lineales nos vemos obligados a tratar dos matrices. Una de ellas se compone de los coeficientes de las incógnitas y se llama *matriz del sistema*. Esta matriz tiene la forma siguiente

$$\begin{pmatrix} a_{11}a_{12}\ldots a_{1m} \\ a_{21}a_{22}\ldots a_{2m} \\ \ldots\ldots\ldots\ldots \\ a_{k1}a_{k2}\ldots a_{km} \end{pmatrix}.$$

La otra se obtiene de la primera por adición de una columna de los segundos miembros

$$\begin{pmatrix} a_{11}a_{12}\ldots a_{1m}b_1 \\ a_{21}a_{22}\ldots a_{2m}b_2 \\ \ldots\ldots\ldots \\ a_{k1}a_{k2}\ldots a_{km}b_k \end{pmatrix}$$

y se denomina *matriz ampliada del sistema*. Observemos, en particular, que el rango de la matriz del sistema coincide con el del sistema de vectores (48.1).

TEOREMA 48.1. (*de Kronecker—Capelli*). *Para que un sistema de ecuaciones algebraicas lineales sea compatible, es necesario y suficiente que el rango de la matriz ampliada del sistema sea igual al rango de la matriz del sistema.*

DEMOSTRACIÓN. Haremos uso de las designaciones aceptadas en el párrafo 22. Los vectores $a_1, a_2, \ldots, a_m, b$ representan en sí, salvo la disposición, las columnas de las matrices en consideración. Puesto que el rango de la matriz coincide con el del sistema de sus vectores columna, entonces para demostrar el teorema basta probar que el sistema será compatible cuando, y sólo cuando, el rango del sistema $a_1, a_2, \ldots, a_m$ coincide con el del sistema $a_1, a_2, \ldots, a_m, b$.

Supongamos que el sistema (48.2) es compatible. Esto significa que la igualdad (22.1) se verifica para cierto surtido de números $z_1, z_2, \ldots, z_m$, es decir, el vector b es una combinación lineal de vectores $a_1, a_2, \ldots, a_m$. Mas, de aquí se deduce que cualquier base del sistema $a_1, a_2, \ldots, a_m$ será, a la vez, la base del sistema $a_1, a_2, \ldots, a_m, b$, es decir, los rangos de ambos sistemas coinciden.

Supongamos ahora que los rangos de estos sistemas coinciden. Elijamos una base cualquiera de $a_1, a_2, \ldots, a_m$. Esta será también la base del sistema $a_1, a_2, \ldots, a_m, b$. Por consiguiente, el vector b se expresa linealmente en términos de una parte de los vectores del sistema $a_1, a_2, \ldots, a_m$. Ya que puede también representarse en forma de una combinación lineal de todos los vectores $a_1, a_2, \ldots$ $\ldots, a_m$, esto significa la compatibilidad del sistema (48.2). El teorema queda demostrado.

El sistema de ecuaciones algebraicas lineales (48.2) se llama *no homogéneo*, si no todos los miembros segundos son nulos. En el caso contrario, se llama *homogéneo*. Todo sistema homogéneo es siempre compatible, puesto que una de sus soluciones es $z_1 = z_2 = \ldots$ $\ldots = z_m = 0$. Un sistema, obtenido del sistema (48.2) como resultado de sustituir todos los segundos miembros por ceros, se denomina sistema homogéneo *reducido*. Si el sistema (48.2) es compatible, cada solución de éste se llamará solución *particular*. El conjunto de todas las soluciones particulares se denominará *solución general del sistema*.

Sirviéndonos de la información obtenida anteriormente, acerca de los planos y sistemas (46.6), (46.7), (46.9) que describen la intersección de los hiperplanos, podemos hacer toda una serie de deducciones referentes a la solución general del sistema de ecuaciones algebraicas lineales. A saber,

La solución general de un sistema homogéneo reducido forma en el espacio P_m un subespacio de dimensión $m - r$, donde r es el rango de la matriz del sistema. Cualquier base de este subespacio lleva el nombre de *sistema fundamental de soluciones*.

La solución general de un sistema no homogéneo es un plano en el espacio P_m, obtenido por desplazamiento de la solución general del sistema homogéneo reducido a cualquier solución particular del sistema no homogéneo.

La diferencia entre cualesquiera dos soluciones particulares de un sistema no homogéneo es una solución particular del sistema homogéneo reducido.

Entre las soluciones particulares de un sistema no homogéneo hay una única solución que es ortogonal a todas las soluciones del sistema homogéneo reducido. Esta solución se llama *normal*.

Para que un sistema compatible tenga una única solución, es necesario y suficiente que el rango de la matriz del sistema sea igual al número de incógnitas.

Para que un sistema homogéneo tenga una solución no nula, es necesario y suficiente que el rango de la matriz del sistema sea inferior al número de incógnitas.

En la investigación realizada de los sistemas de ecuaciones algebraicas lineales el concepto de determinante se ha usado sólo de manera indirecta, principalmente, por intermedio del concepto de rango de la matriz. Sin embargo, en la teoría de los sistemas de ecuaciones el determinante desempeña un papel considerablemente mayor.

Supongamos que la matriz de un sistema es cuadrada. Para que el rango de la matriz del sistema sea inferior al número de incógnitas, es necesario y suficiente que el determinante del sistema sea igual a cero. Por ello,

Un sistema homogéneo tiene una solución no nula cuando, y sólo cuando, el determinante del sistema es igual a cero.

Supongamos ahora que el determinante del sistema es distinto de cero. Esto es indicio de que el rango de la matriz del sistema es m. El rango de la matriz ampliada no puede ser inferior a m. Pero tampoco puede ser superior a m, puesto que no hay menores de orden $m + 1$. Por consiguiente, los rangos de ambas matrices son idénticos, es decir, el sistema en este caso es obligatoriamente compatible. Más aún, tiene una solución única. De este modo,

Si el determinante de un sistema es distinto de cero, el sistema siempre tiene solución y esta solución es única.

Desde el punto de vista de la investigación de la intersección de hiperplanos, a este hecho se le puede atribuir la siguiente forma:

Si los vectores normales de unos hiperplanos forman la base de un espacio, la intersección de dichos hiperplanos no es vacía y contiene sólo un vector.

Indiquemos con d el determinante de un sistema y con d_j, un determinante que se distingue de d sólo en que la j-ésima columna en este último se ha sustituido por la columna de los segundos miembros $b_1, b_2, \ldots, b_m$. En este caso la única solución del sistema puede calcularse según las fórmulas

$$x_j = \frac{d_j}{d}, \quad j = 1, 2, \ldots, n. \qquad (48.3)$$

En efecto, designemos mediante A_{ij} el complemento algebraico del elemento a_{ij} del determinante del sistema. Desarrollando d_j por los elementos de la j-ésima columna, obtenemos

$$d_j = \sum_{i=1}^{n} b_i A_{ij}$$

Sustituyamos ahora las expresiones (48.3) en la k-ésima ecuación arbitraria del sistema

$$\sum_{j=1}^{n} a_{kj}\frac{d_j}{d}=\frac{1}{d}\sum_{j=1}^{n} a_{kj}\sum_{i=1}^{n} b_i A_{ij}=\frac{1}{d}\sum_{j=1}^{n}\sum_{i=1}^{n} a_{kj}b_i A_{ij}=$$

$$=\frac{1}{d}\sum_{i=1}^{n}\sum_{j=1}^{n} b_i a_{kj}A_{ij}=\frac{1}{d}\sum_{i=1}^{n} b_i\sum_{j=1}^{n} a_{kj}A_{ij}=b_k.$$

La suma interior en la última igualdad es igual, de acuerdo con (40.5), (40.8), o bien a d o bien a cero, lo que depende de si $i = k$ ó $i \neq k$.

De este modo, las fórmulas (48.3) proporcionan una expresión explícita para la solución del sistema en términos de sus elementos. En virtud de la unicidad, no existe ninguna otra solución. Estas *fórmulas* llevan el nombre *de Cramer*.

Desde el punto de vista formal, basándose en los cálculos de los determinantes, se pueden resolver cualesquiera sistemas de ecuaciones del tipo (48.2). Al *principio*, calculando varios menores de la matriz del sistema y de la matriz ampliada, comprobamos la compatibilidad del sistema. Supongamos que el sistema resulta ser compatible y el rango de ambas matrices es igual a r. Sin restringir la generalidad, se puede considerar que el menor principal de orden r de la matriz del sistema es distinto de cero. De acuerdo con el teorema 41.1, las últimas $k - r$ filas de la matriz ampliada son combinaciones lineales de sus r primeras filas. Por consiguiente, el sistema

$$\begin{array}{|l|l}
\hline
a_{11}z_1 + \ldots + a_{1r}z_r & + a_{1,r+1}z_{r+1} + \ldots + a_{1m}z_m = b_1, \\
a_{21}z_1 + \ldots + a_{2r}z_r & + a_{2,r+1}z_{r+1} + \ldots + a_{2m}z_m = b_2, \\
\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_{r1}z_1 + \ldots + a_{rr}z_r & + a_{r,r+1}z_{r+1} + \ldots + a_{rm}z_m = b_r \\
\hline
\end{array} \qquad (48.4)$$

es equivalente al sistema (48.2).

Las incógnitas $z_{r+1}, \ldots, z_m$ se llamarán, como antes, *independientes*. Al atribuirles cualesquiera valores, se pueden determinar todas las demás incógnitas, resolviendo el sistema con la matriz del menor principal según, por ejemplo, las fórmulas de Cramer.

Este método de resolver el sistema posee cierto valor sólo en las investigaciones *teóricas*. En cuanto al aspecto *práctico*, resulta mucho más ventajoso servirse del método de Gauss, descrito en el § 22.

Ejercicios.

1. Demuéstrese que la solución general es un conjunto convexo.

2. Demuéstrese que entre todas las soluciones particulares de un sistema no homogéneo la solución normal es de longitud mínima.

3. Demuéstrese que el sistema fundamental es una totalidad de cualesquiera $m - r$ soluciones del sistema homogéneo reducido, para las cuales el determinante compuesto de los valores de las incógnitas independientes es distinto de cero.

4. En el § 22 se ha observado que las pequeñas variaciones en las coordenadas pueden conducir a la violación de la dependencia o independencia lineal de los vectores. ¿Qué deducciones pueden hacerse de este hecho con relación a los problemas referentes a la intersección de hiperplanos?

CAPITULO 6 LÍMITE EN EL ESPACIO LINEAL

§ 49. Espacio métrico

Uno de los conceptos fundamentales del análisis matemático lo constituye el concepto de límite. Está basado en que para los puntos de un eje numérico se ha definido la noción de "cercanía" o, con más precisión, de distancia entre los puntos.

La comparación de las "cercanías" se puede introducir también en los conjuntos cuya naturaleza es totalmente diferente. En el § 29 ya hemos definido la distancia entre los vectores de los espacios lineales dotados de producto escalar, descubriéndose que la distancia citada posee las mismas propiedades (29.5) que tiene la distancia entre los puntos de un eje numérico. La distancia entre los vectores se definía mediante el producto escalar que, a su vez, se introducía axiomáticamente.

Parece natural tratar de introducir *axiomáticamente* la propia distancia, al exigir que se cumplan sin falta las propiedades (29.5).

Ha de ser notado que muchos hechos fundamentales de la teoría del límite en el análisis matemático no están relacionados con el hecho que para los números están definidas unas operaciones algebraicas. Por eso empezaremos por la extensión del concepto de distancia a unos conjuntos arbitrarios de elementos que no son forzosamente vectores de un espacio lineal.

Un conjunto se denomina *espacio métrico*, si a todo par de sus elementos se le ha puesto en correspondencia un número real no negativo, llamado *distancia*, con la particularidad de que se cumplen los axiomas siguientes:

1) $\rho(x, y) = \rho(y, x)$.
2) $\rho(x, y) > 0$, si $x \neq y$; $\rho(x, y) = 0$, si $x = y$.
3) $\rho(x, y) \leqslant \rho(x, z) + \rho(z, y)$

para cualesquiera elementos x, y, z. Estos axiomas se llaman *axiomas de la métrica*; el primero de ellos se denomina *axioma de simetría* y el tercero, *axioma triangular*.

Todo conjunto de elementos en el que se ha definido una relación de igualdad puede ser convertido formalmente en un espacio métrico

al poner

$$\rho(x, y)=\begin{cases}0, & \text{si } x=y,\\ 1, & \text{si } x\neq y.\end{cases} \qquad (49.1)$$

Es fácil comprobar que todos los axiomas de la métrica están cumplidos.

El vector x_0 del espacio métrico X se llama *límite* de la sucesión $\{x_n\}$ de elementos $x_1, x_2, \ldots, x_n, \ldots$ de X, si la sucesión de distancias $\rho(x_0, x_1), \rho(x_0, x_2), \ldots, \rho(x_0, x_n), \ldots$ converge a cero. En este caso suele escribirse

$$x_n \to x_0$$

o bien

$$\lim_{n\to\infty} x_n = x_0$$

y decirse que la sucesión $\{x_n\}$ se llama *convergente en* X o simplemente *convergente*.

Observemos que una misma sucesión de elementos de un mismo conjunto X puede ser convergente o divergente, según sea la métrica introducida en X. Supongamos, por ejemplo, que en un espacio métrico X se ha elegido una sucesión convergente $\{x_n\}$, compuesta por unos elementos distintos dos a dos. Cambiemos la métrica en X, al introducirla de acuerdo con (49.1). En este caso la sucesión $\{x_n\}$ ya no será convergente. Efectivamente, supongamos que $x_n \to x_0'$, es decir, $\rho(x_n, x_0') \to 0$. Con la métrica nueva esto es posible sólo en el caso en que todos los elementos de $\{x_n\}$, a excepción de un número finito, coincidan con x_0'. La contradicción obtenida confirma la afirmación enunciada.

Las dos propiedades que siguen son comunes para cualesquiera sucesiones convergentes.

Si una sucesión $\{x_n\}$ converge, será convergente y tendrá el mismo límite cualquiera de sus subsucesiones. La sucesión puede tener no más que un límite.

La primera propiedad es obvia. Supongamos que la sucesión $\{x_n\}$ tiene dos límites, x_0 e y_0. En este caso, para cualquier número $\varepsilon > 0$, tan pequeño como se quiera, se puede elegir un número N tal que

$$\rho(x_0, x_n) < \frac{\varepsilon}{2}, \qquad \rho(y_0, x_n) < \frac{\varepsilon}{2}$$

para todo $n > N$. De aquí, haciendo uso del axioma triangular, hallamos

$$\rho(x_0, y_0) \leqslant \rho(x_0, x_n) + \rho(x_n, y_0) < \varepsilon.$$

Por ser ε arbitrario, esto significa que $\rho(x_0, y_0) = 0$, es decir, $x_0 = y_0$.

Se denomina *bola* $S(a, r)$ en el espacio métrico X el conjunto de todos los elementos $x \in X$ que satisfacen la condición

$$\rho(a, x) < r. \quad (49.2)$$

El elemento a se llama *centro* de la bola; el número r, *radio* de la bola. Toda bola con el centro en a llamemos *entorno* *) del elemento a. Un conjunto de elementos se denomina *limitado*, si pertenece íntegramente a cierta bola.

Es fácil ver que el elemento x_0 es el límite de la sucesión $\{x_n\}$ cuando, y sólo cuando, cualquier entorno del elemento x_0 contiene todos los elementos de la sucesión que se examina, a partir de cierto número.

En el espacio métrico pueden ser introducidos también muchos otros conceptos de importancia, con los cuales nos encontramos en los conjuntos numéricos. Así por ejemplo, dado el conjunto $M \subset X$, el elemento $x \in X$ se llamará *punto límite* de este conjunto, si cualquier entorno del elemento x contiene al menos un solo elemento del conjunto M que no coincide con x. Un conjunto obtenido por agregación a M de todos sus puntos límite se denomina *adherencia* del conjunto M y se indica con $\overline{M}$. El conjunto M se llama *cerrado*, si $M = \overline{M}$.

Examinemos los puntos límite de la bola (49.2). Probemos que todos ellos satisfacen la condición

$$\rho(a, x) \leqslant r. \quad (49.3)$$

Efectivamente, supongamos que existe por lo menos un punto límite x' para la bola (49.2), para el cual $\rho(a, x') > r$. Por definición de punto límite, todo entorno de elemento x' debe contener al menos un elemento de la bola (49.2) que no sea coincidente con x'. Pero el entorno cuyo radio es $0,5(\rho(a, x') - r)$ no contiene, a ciencia cierta, ninguno de los elementos de tal índole. De conformidad con lo dicho:

El conjunto $\overline{S}(a, r)$ de todos los elementos x que satisfacen la condición (49.3) se llama bola cerrada.

Ejercicios.

1. Demuéstrese que si $x_n \to x$, entonces $\rho(x_n, a) \to \rho(x, a)$ para todo elemento x.

2. ¿Será el conjunto de todos los números reales un espacio métrico, si para los números x, y ponemos

$$\rho(x, y) = \operatorname{arc\,tg} |x - y|?$$

3. ¿Podrá un conjunto compuesto por un número finito de elementos tener puntos límite?

---

*) Se usa también el término *vecindad*. (*N. del Tr.*)

§ 50. Espacio completo

La sucesión $\{x_n\}$ de elementos de un espacio métrico se llama *fundamental* o *convergente en sí*, si para cualquier número $\varepsilon > 0$ existe tal número N que $\rho(x_n, x_m) < \varepsilon$ para $n, m > N$.

Toda sucesión fundamental es limitada. En efecto, siendo ε dado, elijamos un número N, de acuerdo con la definición, y tomemos un número arbitrario $n_0 > N$. Todos los elementos de la sucesión, a partir de x_{n_0}, pertenecen, a ciencia cierta, a una bola de radio ε y centro x_{n_0}. Mientras tanto todos los elementos pertenecen a una bola cuyos centro y radio, respectivamente, son x_{n_0} y el máximo de los números

$$\varepsilon, \ \rho(x_1, x_{n_0}), \ \ldots, \ \rho(x_{n_0-1}, x_{n_0}).$$

Si la sucesión es convergente, será fundamental. Supongamos que la sucesión $\{x_n\}$ converge a x_0. Entonces para todo $\varepsilon > 0$ existe N tal que

$$\rho(x_n, x_0) < \frac{\varepsilon}{2}$$

para $n > N$. De acuerdo con el axioma triangular,

$$\rho(x_n, x_m) \leqslant \rho(x_n, x_0) + \rho(x_0, x_m) < \varepsilon$$

para $n, m > N$, lo que significa precisamente el carácter fundamental de la sucesión $\{x_n\}$.

Para el conjunto de todos los números reales es cierta también la afirmación recíproca. A saber, toda sucesión fundamental es convergente. No obstante, en el caso general esto ya no es cierto, lo que se confirma con el ejemplo de un espacio métrico del cual se ha eliminado por lo menos un solo punto límite.

Un espacio métrico se denomina *completo*, si toda sucesión fundamental en él es convergente.

En los espacios métricos completos tiene lugar un teorema que es análogo al teorema sobre segmentos encajados para los números reales. Sea dada una sucesión de bolas. Llamaremos estas bolas *encajadas* una dentro de la otra, si toda bola consecutiva está contenida dentro de la anterior.

TEOREMA 50.1 *Supongamos que en el espacio métrico completo X sea dada la sucesión $\{\overline{S}(a_n, \varepsilon_n)\}$ de bolas cerradas encajadas una dentro de la otra. Si la sucesión de radios tiende hacia cero, existe un único elemento de X perteneciente a todas estas bolas.*

DEMOSTRACIÓN. Examinemos la sucesión $\{a_n\}$. Puesto que $\overline{S}(a_{n+p}, \varepsilon_{n+p}) \subset \overline{S}_r(a_n, \varepsilon_n)$ para cualquier $p \geqslant 0$, entonces $a_{n+p} \in$

$\in \overline{S}(a_n, \varepsilon_n)$. Por consiguiente,

$$\rho(a_{n+p}, a_n) \leqslant \varepsilon_n,$$

de donde se desprende que la sucesión $\{a_n\}$ es fundamental.

El espacio X es completo y, por ende, la sucesión $\{a_n\}$ converge a cierto límite a de X. Tomemos una bola cualquiera $\overline{S}(a_k, \varepsilon_k)$. A esta bola le pertenecen todos los términos de la sucesión $\{a_n\}$, a partir de a_k. En vista de que las bolas son cerradas, el límite de esta sucesión también pertenece a $\overline{S}(a_k, \varepsilon_k)$. De este modo, a pertenece a todas las bolas.

Admitamos luego que existe otro elemento, b, perteneciente también a todas las bolas. De acuerdo con el axioma triangular

$$\rho(a, b) \leqslant \rho(a, a_n) + \rho(a_n, b) \leqslant 2\varepsilon_n.$$

Puesto que ε_n puede ser tan pequeño como se quiera, esto significa que $\rho(a, b) = 0$, es decir, $a = b$.

Como ejemplos más importantes de espacios completos sirven los conjuntos de números reales y complejos. En este caso se supone que la distancia entre los números coincide con el módulo de la diferencia entre ellos. La completitud del conjunto de números reales se demuestra en el curso del análisis matemático. Mostremos la completitud del conjunto de números complejos.

Convengamos en considerar que los números complejos vienen dados en forma algebraica. La distancia entre los números

$$u = a + ib, \qquad v = c + id$$

se introducirá de conformidad con la regla

$$\rho(u, v) = |u - v|, \tag{50.1}$$

donde

$$|u - v|^2 = (a - c)^2 + (b - d)^2. \tag{50.2}$$

Es evidente que los axiomas de la métrica están cumplidos.

Consideremos una sucesión $\{z_k = a_k + ib_k\}$ de números complejos. Supongamos que es fundamental. Siendo $\varepsilon > 0$, hallamos tal N que para cualesquiera $n, m > N$ sea

$$|z_n - z_m| < \varepsilon.$$

De (50.2) se infiere que en este caso

$$|a_n - a_m| < \varepsilon, \qquad |b_n - b_m| < \varepsilon, \tag{50.3}$$

es decir, las sucesiones $\{a_k\}$ y $\{b_k\}$ son también fundamentales. Dado que el conjunto de números reales es completo, existen unos números a, b para los cuales

$$a_k \to a, \qquad b_k \to b.$$

Pasando al límite en las desigualdades (50.3), obtendremos

$$|a_n - a| \leq \varepsilon, \qquad |b_n - b| \leq \varepsilon.$$

Al designar

$$z = a + ib,$$

encontramos que

$$\rho(z_n, z) \leq \sqrt{2}\,\varepsilon$$

para todo $n > N$. Mas, esto significa que la sucesión fundamental $\{z_k\}$ es convergente.

Observemos, a título de corolario, que la sucesión $\{z_k = a_k + ib_k\}$ converge al número $z = a + ib$ cuando, y sólo cuando, las sucesiones $\{a_k\}$ y $\{b_k\}$ convergen a los números a y b, respectivamente.

El espacio completo de números complejos tiene mucho en común con el espacio de números reales. En particular, toda sucesión acotada de números complejos tiene una subsucesión convergente. En efecto, esta afirmación es verídica para toda sucesión acotada de números reales. Luego, es evidente que si es acotada la sucesión $\{z_k = a_k + ib_k\}$, lo serán también las sucesiones $\{a_k\}$ y $\{b_k\}$. Puesto que la sucesión $\{a_k\}$ es acotada, tiene subsucesión convergente $\{a_{\nu_k}\}$. Consideremos la sucesión $\{b_{\nu_k}\}$. Es acotada y por esta razón también tiene subsucesión convergente $\{b_{\nu_{k_n}}\}$. Está claro que $\{a_{\nu_{k_n}}\}$ será convergente. Por consiguiente, la subsucesión $\{z_{\nu_{k_n}}\}$ será también convergente.

Por analogía con el espacio real, en el espacio complejo se introduce la noción de sucesión infinita creciente. A saber, la sucesión $\{z_k\}$ se denomina *infinita creciente*, si para un número A, tan grande como se quiera, se puede indicar un número N tal que para todo $k > N$ se cumpla la desigualdad $|z_k| > A$. Es evidente que en toda sucesión no acotada siempre puede elegirse una subsucesión infinita creciente.

Ejercicios.

1. ¿Será un espacio completo el conjunto de todos los números reales, si para los números x, y hacemos

$$\rho(x, y) = \operatorname{arctg}|x - y|?$$

2. Demuéstrese que todo conjunto cerrado de un espacio completo es por sí mismo un espacio completo.
3. ¿Será necesariamente un espacio completo todo conjunto cerrado de un espacio métrico arbitrario?
4. Constrúyase una métrica, para la cual el conjunto de todos los números complejos no sea un espacio completo.

§ 51. Desigualdades auxiliares

Obtengamos algunas desigualdades que se emplearán en las investigaciones inmediatas. Tomemos un número positivo arbitrario α y examinemos una función exponencial $y = \alpha^x$ (fig. 51.1). Sean x_1, x_2 dos números reales distintos. Tracemos una recta por los puntos cuyas coordenadas son (x_1, α^{x_1}), (x_2, α^{x_2}). Teniendo presentes las propiedades de una función exponencial, concluimos que con el cambio de argumento en el segmento $[x_1, x_2]$ todos los puntos de la misma no serán superiores a los puntos de la recta construida.

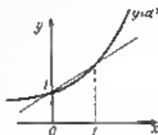


Fig. 51.1.

Ahora, sea $x_1 = 0$ y sea $x_2 = 1$. En este caso la ecuación de la recta a examinar será $y = \alpha x + (1 - x)$. Por consiguiente,

$$\alpha^x \leqslant \alpha x + (1 - x) \tag{51.1}$$

para $0 \leqslant x \leqslant 1$.

Llamemos *conjugados* los números positivos p, q, si satisfacen la correlación

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{51.2}$$

Está claro que p, $q > 1$.

Para cualesquiera números positivos a, b el número $a^p b^{-q}$ será también positivo y se la puede tomar en calidad de α de (51.1). Si se considera que $x = p^{-1}$, entonces $1 - x = q^{-1}$. Ahora, de (51.1) se deduce la validez de la desigualdad siguiente:

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q} \tag{51.3}$$

para cualesquiera a, b positivos y p, q conjugados. Evidentemente, esta desigualdad tiene lugar para todos los números no negativos a, b.

Consideraremos dos vectores arbitrarios x, y, pertenecientes al espacio $\mathbf{R}_n$ o $\mathbf{C}_n$. Supongamos que estos vectores vienen dados mediante sus coordenadas

$$x = (x_1, x_2, \ldots, x_n),$$
$$y = (y_1, y_2, \ldots, y_n).$$

Demos a conocer la así llamada *desigualdad de Hölder*

$$\sum_{k=1}^{n} |x_k y_k| \leqslant \Big(\sum_{k=1}^{n} |x_k|^p\Big)^{1/p} \Big(\sum_{k=1}^{n} |y_k|^q\Big)^{1/q}. \tag{51.4}$$

Hemos de notar que si entre los vectores x, y hay por lo menos un vector nulo, la desigualdad de Hölder es, evidentemente, lícita. Por esto podemos considerar que $x \neq 0$, $y \neq 0$. Supongamos que la desigualdad se cumpla para cualesquiera vectores no nulos x, y. En este caso se cumple también para los vectores λx, μy con cualesquiera λ, μ. Por eso es suficiente demostrarla sólo para el caso en que

$$\sum_{k=1}^{n} |x_k|^p = \sum_{k=1}^{n} |y_k|^q = 1. \qquad (51.5)$$

Haciendo ahora $a = |x_k|$, $b = |y_k|$ en la desigualdad (51.3) y sumando según k desde 1 hasta n, obtendremos, tomando en consideración (51.2), (51.5):

$$\sum_{k=1}^{n} |x_k y_k| \leqslant 1.$$

Esto es precisamente la desigualdad de Hölder para el caso (51.5).

Pasemos ahora a la demostración de *la desigualdad de Minkowski* en la que para cualesquiera vectores x, y de $\mathbf{R}_n$ o $\mathbf{C}_n$ se verifica

$$\Big(\sum_{k=1}^{n} |x_k + y_k|^p\Big)^{1/p} \leqslant \Big(\sum_{k=1}^{n} |x_k|^p\Big)^{1/p} + \Big(\sum_{k=1}^{n} |y_k|^p\Big)^{1/p} \qquad (51.6)$$

para todo $p \geqslant 1$.

La desigualdad de Minkowski es obvia para $p = 1$, puesto que el módulo de una suma de dos números no es superior a la suma de sus módulos. Además, se cumple a ciencia cierta, si por lo menos uno de los vectores x, y es nulo. Por ello podemos limitarnos al examen del caso en que $p > 1$ y $x \neq 0$. Escribamos la identidad

$$(|a| + |b|)^p = (|a| + |b|)^{p-1} |a| + (|a| + |b|)^{p-1} |b|.$$

Al poner en ésta $a = x_k$, $b = y_k$ y al sumar según k desde 1 hasta n, obtenemos

$$\sum_{k=1}^{n} (|x_k| + |y_k|)^p = \sum_{k=1}^{n} (|x_k| + |y_k|)^{p-1} |x_k| + \sum_{k=1}^{n} (|x_k| + |y_k|)^{p-1} |y_k|.$$

Apliquemos la desigualdad de Hölder a cada una de las dos sumas que figuran en el segundo miembro de esta correlación. Teniendo presente que $(p-1)\,q = p$, obtenemos

$$\sum_{k=1}^{n} (|x_k| + |y_k|)^p \leqslant \Big(\sum_{k=1}^{n} (|x_k| + |y_k|)^p\Big)^{1/q} \times$$
$$\times \Big(\Big(\sum_{k=1}^{n} |x_k|^p\Big)^{1/p} + \Big(\sum_{k=1}^{n} |y_k|^p\Big)^{1/p}\Big)$$

Al dividir ambos miembros de la desigualdad por

$$\Big(\sum_{n=1}^{n}(|x_n|+|y_n|)^p\Big)^{1/q},$$

encontramos que

$$\Big(\sum_{n=1}^{n}(|x_n|+|y_n|)^p\Big)^{1/p}\leqslant\Big(\sum_{n=1}^{n}|x_n|^p\Big)^{1/p}+\Big(\sum_{n=1}^{n}|y_n|^p\Big)^{1/p},$$

de donde proviene inmediatamente la desigualdad (51.6).

Ejercicios.

1. Dedúzcase la desigualdad de Cauchy—Buniakovski a partir de la desigualdad de Hölder.
2. Estúdiese la desigualdad de Hölder para $n\to\infty$.
3. Estúdiese la desigualdad de Minkowski para $n\to\infty$.

§ 52. Espacio normalizado

Al concepto de espacio métrico hemos llegado concentrando nuestra atención sólo en una propiedad del conjunto, a saber, en la presencia de una distancia en dicho conjunto. De modo análogo, al concentrar nuestra atención en las operaciones en un conjunto, llegamos al concepto de espacio lineal. Ahora consideraremos los espacios lineales provistos de una métrica.

Evidentemente, si el concepto de distancia no está relacionado de tal o cual modo con las operaciones sobre elementos, resulta imposible construir una teoría enjundiosa cuyos hechos unan juntar los conceptos algebraicos y métricos. Por esta razón impondremos sobre la métrica, introducida en el espacio lineal, unas condiciones complementarias.

Ya nos hemos encontrado en realidad con los espacios lineales métricos. Son, por ejemplo, los espacios euclídeo y unitario con una métrica (28.4). No obstante, la necesidad en tal métrica no surge siempre. La introducción de un producto escalar significa, de hecho, que se introduce no sólo la distancia entre los elementos, sino también los ángulos entre los mismos. Con mayor frecuencia se requiere que en el espacio lineal sea dado sólo el concepto aceptable de distancia. Los más importantes espacios lineales de este tipo son los así llamados *espacios normalizados*.

Un espacio lineal X, sea real o complejo, se llama normalizado, si a cada vector $x\in X$ se le ha puesto en correspondencia un número real $\|x\|$, llamado *norma* del vector x, con la particularidad de que se consideren cumplidos los siguientes axiomas:

1) $\|x\|>0$, si $x\neq 0$, $\|0\|=0$,
2) $\|\lambda x\|=|\lambda|\,\|x\|$,
3) $\|x+y\|\leqslant\|x\|+\|y\|$

para cualesquiera vectores x, y y todo número λ. El segundo axioma lleva el nombre de *axioma de homogeneidad absoluta de la norma*, el tercer axioma se llama *axioma de la desigualdad triangular*.

Del axioma de homogeneidad absoluta de la norma proviene que para todo vector x no nulo se pueda encontrar un número λ tal que la norma del vector λx sea igual a uno. Para ello será suficiente tomar $\lambda = \|x\|^{-1}$. Un vector cuya norma es igual a la unidad se llamará *normalizado*.

De la desigualdad triangular para las normas se desprende una correlación muy útil. Tenemos $\|x\| \leqslant \|y\| + \|x - y\|$ para cualesquiera x, y. Por consiguiente, $\|x\| - \|y\| \leqslant \|x - y\|$. Intercambiando x, y de lugares, obtenemos $\|y\| - \|x\| \leqslant \leqslant \|x - y\|$. Por esto

$$| \|x\| - \|y\| | \leqslant \|x - y\|. \tag{52.1}$$

Un espacio normalizado se convierte con facilidad en un espacio métrico, si ponemos

$$\rho(x, y) = \|x - y\|. \tag{52.2}$$

En efecto, $\rho(x, y) = 0$ quiere decir que $\|x - y\| = 0$, lo que de acuerdo con el axioma 1, significa $x = y$. La simetría de la distancia introducida es obvia. Por fin, la desigualdad triangular para la distancia es simplemente un corolario de la desigualdad triangular para la norma. A saber,

$$\rho(x, y) = \|x - y\| = \|(x - z) + (z - y)\| \leqslant$$
$$\leqslant \|x - z\| + \|z - y\| = \rho(x, z) + \rho(z, y).$$

Observemos que

$$\|x\| = \rho(x, 0). \tag{52.3}$$

La métrica (52.2) definida en el espacio lineal posee, además, las siguientes propiedades:

$$\rho(x + z, y + z) = \rho(x, y)$$

para cualesquiera x, y, $z \in X$, es decir, la distancia no varía con el desplazamiento de los vectores, y

$$\rho(\lambda x, \lambda y) = |\lambda| \rho(x, y)$$

para cualesquiera vectores x, $y \in X$ y todo número λ, es decir, la distancia es una función absolutamente homogénea.

Si en el espacio lineal métrico X una métrica, cualquiera que sea, satisface estas dos exigencias complementarias, entonces X puede considerarse como espacio normalizado, siempre que la norma se defina mediante la igualdad (52.3) para todo $x \in X$.

Tomando en consideración las correlaciones del § 29, es fácil establecer que *un espacio lineal provisto de un producto escalar es un*

espacio normalizado. En este caso por norma del vector conviene tomar su longitud.

Se pueden aducir también otros ejemplos de introducción de la norma. Supongamos que en un espacio lineal los vectores vienen dados por medio de sus coordenadas respecto de cierta base. Si $x = (x_1, x_2, \dots, x_n)$, ponemos

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} \qquad (52.4)$$

donde $p \geq 1$. El cumplimiento de los primeros dos axiomas para la norma es obvio, mientras que el cumplimiento del tercer axioma se deduce de la desigualdad de Minkowski (51.6). Son de mayor uso las siguientes normas:

$$\|x\|_1 = \sum_{k=1}^{n} |x_k|, \quad \|x\|_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2},$$
$$\|x\|_\infty = \max_{1 \leq k \leq n} |x_k|. \qquad (52.5)$$

La segunda de dichas normas se llama, a menudo, norma *euclídea* y se indica con $\|x\|_E$.

En adelante se considerarán sólo espacios normalizados los que contengan la métrica (52.2). La convergencia de una sucesión de vectores en tal métrica la llamaremos *convergencia en norma* y la acotación de un conjunto de vectores, *acotación en norma*, etc.

Ejercicios.

1. Demuéstrese que existe una sucesión de vectores cuyas normas forman una sucesión infinita creciente.

2. Demuéstrese que para cualesquiera números λ_i y vectores e_i se verifica la desigualdad

$$\left\|\sum_{i=1}^{n} \lambda_i e_i\right\| \leq \sum_{i=1}^{n} |\lambda_i| \, \|e_i\|.$$

3. Demuéstrese que si $x_n \to x$, $y_n \to y$, $\lambda_n \to \lambda$, entonces

$$\|x_n\| \to \|x\|, \quad x_n + y_n \to x + y, \quad \lambda_n x_n \to \lambda x.$$

§ 53. Convergencia en norma y convergencia coordenada

Un espacio lineal de dimensión finita, real o complejo, además de la convergencia en norma admite también la introducción de otro concepto de convergencia. Consideremos el espacio X y sea $e_1, e_2, \dots, e_n$ su base. Para cualquier suce-

sión $\{x_m\}$ de los vectores de X existen los desarrollos

$$x_m = \sum_{k=1}^{n} \xi_k^{(m)} e_k. \tag{53.1}$$

Si para el vector

$$x_0 = \sum_{k=1}^{n} \xi_k^{(0)} e_k \tag{53.2}$$

tienen lugar las correlaciones límites

$$\lim_{m \to \infty} \xi_k^{(m)} = \xi_k^{(0)} \tag{53.3}$$

para cualesquiera $k = 1, 2, \ldots, n$, entonces diremos que tiene lugar la convergencia *coordenada* de la sucesión $\{x_m\}$ al vector x_0.

La convergencia coordenada es bien natural. Si dos vectores son "cercanos", se puede suponer que han de ser "cercanas" también las coordenadas correspondientes en el desarrollo según una misma base. Los espacios normalizados de dimensión finita son remarcables por el hecho de que en estos espacios los conceptos de convergencia en norma y de convergencia coordenada son *equivalentes*.

Es fácil probar que de la convergencia coordenada proviene la convergencia en norma. En efecto, supongamos que tienen lugar las correlaciones límites (53.3). Haciendo uso de los axiomas de homogeneidad absoluta de la norma y de la desigualdad triangular, concluimos a base de (53.3), (53.2) que

$$\| x_m - x_0 \| = \Big\| \sum_{k=1}^{n} (\xi_k^{(m)} - \xi_k^{(0)}) e_k \Big\| \leq \sum_{k=1}^{n} |\xi_k^{(m)} - \xi_k^{(0)}| \, \| e_k \| \to 0.$$

La demostración de la afirmación inversa es considerablemente más compleja.

LEMA 53.1. *Si en un espacio normalizado de dimensión finita la sucesión de vectores está acotada en norma, serán acotadas también las sucesiones numéricas de todas las coordenadas en el desarrollo de los vectores según cualquier base.*

DEMOSTRACIÓN. Supongamos que todo vector de la sucesión $\{x_m\}$ está representado en la forma (53.1). Introduzcamos la designación

$$\sigma_m = \sum_{k=1}^{n} |\xi_k^{(m)}|$$

y demostremos que la sucesión $\{\sigma_m\}$ está acotada.

Supóngase que no es así. Entonces, de dicha sucesión se puede elegir una subsucesión infinita creciente $\{\sigma_{m_p}\}$. Pongamos

$$y_p = \frac{1}{\sigma_{m_p}} x_{m_p},$$

Si

$$y_p = \sum_{k=1}^{n} \eta_k^{(p)} e_k,$$

entonces, desde luego

$$\eta_k^{(p)} = \frac{\xi_k^{(m_p)}}{\sigma_{m_p}}$$

para cualesquiera $k = 1, 2, \ldots, n$ y m_p, y concluimos que

$$\sum_{k=1}^{n} |\eta_k^{(p)}| = 1. \qquad (53.4)$$

Las sucesiones $\{\eta_k^{(p)}\}$ son acotadas, puesto que, de conformidad con (53.4), $|\eta_k^{(p)}| \leqslant 1$. Por consiguiente, se puede elegir tal subsucesión de vectores $\{y_{r_1}\}$ que la subsucesión $\{\eta_1^{(r_1)}\}$ será convergente, es decir

$$\lim_{r_1 \to \infty} \eta_1^{(r_1)} = \eta_1$$

para cierto número η_1. En la subsucesión $\{y_{r_1}\}$ puede elegirse, a su vez, una subsucesión $\{y_{r_2}\}$, para la cual

$$\lim_{r_2 \to \infty} \eta_2^{(r_2)} = \eta_2$$

para cierto número η_2. En este caso, como hasta ahora,

$$\lim_{r_2 \to \infty} \eta_1^{(r_2)} = \eta_1.$$

Continuando este proceso, elegimos en la sucesión $\{y_r\}$ tal subsucesión $\{y_{p_n}\}$ que existirán los límites

$$\lim_{p_n \to \infty} \eta_k^{(p_n)} = \eta_k \qquad (53.5)$$

para cualesquiera $k = 1, 2, \ldots, n$. De acuerdo con (53.4),

$$\sum_{k=1}^{n} |\eta_k| = 1 \qquad (53.6)$$

De la convergencia coordenada se desprende la convergencia en norma, por lo cual, las correlaciones límite (53.5) significan que

$$\lim_{p_n \to \infty} \| y_{p_n} - y \| = 0, \qquad (53.7)$$

donde

$$y = \sum_{k=1}^{n} \eta_k e_k.$$

El vector y no debe ser nulo en virtud de (53.6). Por otro lado, se tiene

$$\|y_{p_n}\| = \frac{\|x_{m_{p_n}}\|}{\alpha_{m_{p_n}}} \to 0,$$

puesto que la sucesión $\{x_{m_{p_n}}\}$ está acotada en norma, y la subsucesión $\{\alpha_{m_{p_n}}\}$ es infinita creciente. Por consiguiente de (53.7) se desprende que $\|y\| = 0$, es decir, y es un vector nulo. La contradicción obtenida demuestra la validez de la afirmación del lema.

TEOREMA 53.1. *En el espacio normalizado de dimensión finita, de la convergencia en norma se infiere la convergencia coordenada.*

DEMOSTRACION. Sea dada la sucesión de vectores $\{x_m\}$ que converge en norma al vector x_0. Es evidente que basta examinar el caso en que $x_0 = 0$ y en la sucesión $\{x_m\}$ no haya vectores nulos. Representemos los vectores x_m en forma de los desarrollos (53.1). La sucesión de vectores

$$y_m = \frac{1}{\|x_m\|} x_m$$

será acotada en norma y, de acuerdo con el lema 53.1, deben ser acotadas las sucesiones de números

$$\eta_k^{(m)} = \frac{\xi_k^{(m)}}{\|x_m\|}$$

para cualesquiera $k = 1, 2, \ldots, n$. Puesto que $\|x_m\| \to 0$, esto es posible sólo en el caso en que $\xi_k^{(m)} \to 0$ para todo k. Pero esto es precisamente testimonio de que tiene lugar la convergencia coordenada de la sucesión $\{x_m\}$ al vector x_0.

La convergencia coordenada se emplea con toda la eficacia en las investigaciones *teóricas*, mientras que la convergencia en norma es mucho más cómoda para utilizarla en las aplicaciones *prácticas*. Esto se debe, principalmente, a que en la investigación de los espacios lineales de gran dimensión resulta difícil operar con un elevado número de sucesiones coordenadas. Además, no siempre se conoce aunque sea una sola base. Pero, si incluso sabemos la base, el empleo de la misma nos conduce, en la mayoría de los casos, a cálculos voluminosos y no justificados.

Ejercicios.

1. Cuando se demuestra la equivalencia de dos tipos de convergencia ¿será importante el requisito de que un espacio sea de dimensión finita?
2. Demuéstrese que si un conjunto de vectores de un espacio de dimensión finita está acotado en una norma, estará acotado también en cualquier otra norma.
3. Demuéstrese que si en un espacio de dimensión finita $x_n \to x$ en una norma, entonces $x_n \to x$ en cualquier otra norma.

§ 54. Completitud de los espacios normalizados

Los espacios normalizados de dimensión finita representan espacios en los cuales tienen lugar varias analogías de las afirmaciones relacionadas con el concepto de límite en los conjuntos de números. He aquí algunas de estas analogías.

LEMA 54.1. *De toda sucesión acotada de vectores de un espacio normalizado de dimensión finita se puede elegir una subsucesión convergente en dicho espacio.*

DEMOSTRACIÓN. Sea $\{x_m\}$ una sucesión arbitraria acotada en norma. Representemos los vectores x_m en forma de los desarrollos (53.1). De acuerdo con el lema 53.1, las sucesiones $\{\xi_k^{(m)}\}$ serán acotadas para cualesquiera $k = 1, 2, \ldots, n$. De un modo semejante al usado en la demostración del lema 53.1, elijamos en la sucesión $\{x_m\}$ una subsucesión $\{x_{m_n}\}$, para la cual existen correlaciones límites $\xi_k^{(m_n)} \to \xi_k^{(0)}$ para todo k. De aquí se deduce que la subsucesión $\{x_{m_n}\}$ converge al vector (53.2).

El lema demostrado es una analogía del lema de Bolzano—Weierstrass que es conocido en el curso del análisis matemático. El lema es de mucha importancia en las investigaciones de cualesquiera espacios normalizados de dimensión finita. Ilustrémoslo demostrando algunas afirmaciones.

TEOREMA 54.1. *Todo espacio normalizado de dimensión finita es completo.*

DEMOSTRACIÓN. Sea $\{x_m\}$ una sucesión fundamental. Está acotada. Elijamos en ella una subsucesión convergente $\{x_{m_n}\}$ e indiquemos con x_0 su límite. Entonces

$$\|x_m - x_0\| \leqslant \|x_m - x_{m_n}\| + \|x_{m_n} - x_0\|.$$

Tomemos un número arbitrario $\varepsilon > 0$. Ya que la sucesión $\{x_m\}$ es fundamental, existe un número N_1 tal que $\|x_m - x_{m_n}\| < \varepsilon/2$ cuando $m, m_n > N_1$. En vista de que la sucesión $\{x_{m_n}\}$ converge a x_0, existe un número N_2 tal que $\|x_{m_n} - x_0\| < \varepsilon/2$ cuando $m_n > N_2$. Si N es el máximo de los números N_1, N_2, entonces, siendo $m > N$,

$$\|x_m - x_0\| < \varepsilon.$$

El número ε es arbitrario. Por lo tanto, la sucesión fundamental $\{x_m\}$ converge en norma al vector x_0.

LEMA 54.2. *Todo subespacio de dimensión finita X_0 de un espacio normalizado X es un conjunto cerrado.*

DEMOSTRACIÓN. Consideremos en el espacio normalizado X un subespacio de dimensión finita X_0. Supongamos que el vector $x \in X$ es un punto límite para X_0. Esto quiere decir que existe una sucesión de vectores $\{x_m\}$ de X_0, no coincidentes con x, tal que

$\| x_m - x \| \to 0$. La sucesión $\{x_m\}$ es acotada y, por ende, podemos elegir de la misma una subsucesión $\{x_{m_p}\}$ que, por ser X_0 completo, converge a cierto vector $x_0 \in X_0$. Ahora tenemos

$$\| x - x_0 \| \leqslant \| x - x_{m_p} \| + \| x_{m_p} - x_0 \| \to 0,$$

es decir, $x = x_0$.

LEMA 54.3. *Sea X un espacio normalizado y sea X_0 su subespacio de dimensión finita no coincidente con X. Existe un vector normalizado $x \in X_1$ tal que $\| x - x_0 \| \geqslant 1$ para cualquier vector $x_0 \in X_0$.*

DEMOSTRACIÓN. Como que X_0 no coincide con X, existe un vector $x' \notin X_0$. Puesto que X_0 es cerrado, tenemos

$$\inf_{x_0 \in X_0} \| x' - x_0 \| = d > 0. \qquad (54.1)$$

Por definición de la cota exacta inferior, en X_0 habrá un vector $x_0^{(k)}$, para el cual

$$d \leqslant \| x' - x_0^{(k)} \| \leqslant \frac{d}{1 - 2^{-k}}.$$

La sucesión $\{x_0^{(k)}\}$ es acotada. Elijamos en la misma una subsucesión $\{x_0^{(k_p)}\}$ que converge, en virtud de que X_0 es completo, a cierto vector $x_0' \in X_0$. Para este vector, evidentemente,

$$\| x' - x_0' \| = d. \qquad (54.2)$$

Hagamos

$$x = \frac{1}{d}(x' - x_0').$$

Está claro que $\| x \| = 1$. Además, si $x_0 \in X_0$, entonces, de acuerdo con (54.1), tendremos

$$\| x - x_0 \| = \left\| \frac{1}{d} x' - \frac{1}{d} x_0' - x_0 \right\| = \frac{1}{d} \| x' - (x_0' + dx_0) \| \geqslant \frac{1}{d} d = 1,$$

puesto que el vector $x_0' + dx_0$ pertenece a X_0.

Hemos demostrado al mismo tiempo que la cota inferior (54.1) se logra por lo menos en un vector $x_0' \in X_0$. En la correlación $\| x - x_0 \| \geqslant 1$ la igualdad se consigue a ciencia cierta para $x_0 = 0$.

Observemos, como conclusión, que el lema 54.1 que desempeña un papel tan importante en los espacios de dimensión finita, no es válido en *ningún* espacio de dimensión infinita. A saber, es válido

LEMA 54.4. *Si en cada sucesión acotada de vectores del espacio normalizado X se puede elegir una subsucesión convergente, entonces el espacio X es de dimensión finita.*

DEMOSTRACIÓN. Supongamos lo contrario. Sea X un espacio de dimensión infinita. Elijamos un vector normalizado arbitrario x_1 y designemos con L_1 su cápsula lineal. De acuerdo con el lema 54.3, existe un vector normalizado x_2 tal que $\| x_2 - x_1 \| \geqslant 1$. Designemos mediante L_2 la cápsula lineal de los vectores x_1, x_2. Continuando los razonamientos, hallemos la sucesión $\{x_k\}$ de los vectores nor-

normalizados que satisfacen las desigualdades $\| x_n - x_k \| \geq 1$ para todo $k < n$. Por consiguiente, de esta sucesión no se puede elegir ninguna subsucesión convergente. Esto contradice la hipótesis del lema, por lo cual, la suposición acerca de que el espacio X es de dimensión infinita era errónea.

Ejercicios.

1. Demuéstrese que un plano en el espacio normalizado de dimensión finita es un conjunto cerrado.

2. Demuéstrese que el conjunto de vectores x de un espacio de dimensión finita que satisfacen la condición $\| x \| \leq \alpha$, es un conjunto cerrado.

3. Demuéstrese que en un conjunto acotado cerrado de vectores de un espacio de dimensión finita existen unos vectores en los cuales se logran tanto la cota inferior como la superior de los valores de cualquier norma.

4. Demuéstrese que para cualesquiera dos normas $\| x \|_I$, $\| x \|_{II}$ en un espacio de dimensión finita existen tales números positivos α, β que

$$\alpha \| x \|_I \leq \| x \|_{II} \leq \beta \| x \|_I$$

para todos los vectores x. Los números α, β no dependen de x.

§ 55. Límite y procesos de cálculo

En un espacio métrico completo el concepto de límite es de amplio uso al construir y argumentar los más diversos procesos de cálculo. Consideremos, a título de ejemplo, uno de los métodos de resolución de los sistemas de ecuaciones algebraicas lineales.

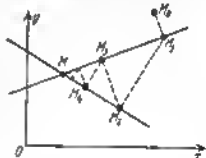


Fig. 55.1.

Sea dado un sistema de dos ecuaciones con dos incógnitas. Convengamos en considerar que el sistema es compatible y tiene una sola solución. Con el fin de simplificar la exposición, supongamos que todos los coeficientes son reales. Cada una de las ecuaciones

$$a_{11}x + a_{12}y = f_1,$$
$$a_{21}x + a_{22}y = f_2$$

del sistema define en el plano una recta. El punto M, donde se cortan estas rectas, determina la solución del sistema (fig. 55.1).

Tomemos un punto arbitrario M_0 que se halla en el plano fuera de las rectas mencionadas. Tracemos desde este punto una perpendicular a cualquiera de las rectas. El pie M_1 de la perpendicular queda más próximo al punto M que el punto M_0, puesto que una proyección es siempre menor que una oblicua. Tracemos a continuación una perpendicular desde el punto M_1 a la otra recta. El pie M_2

de esta última perpendicular se encontrará aún más próximo a la solución. Realizando sucesivamente la proyección, una vez a una recta y otra vez, a la otra, obtendremos una sucesión $\{M_k\}$ de puntos en el plano la cual converge al punto M. Merece subrayar que la convergencia de la sucesión construida tiene lugar para cualquier ubicación inicial del punto M_0.

Este ejemplo sugiere cómo se puede construir el proceso de cálculo para resolver un sistema de ecuaciones algebraicas lineales en la forma general (48.2). Sustituyamos el problema dado por uno equivalente que consiste en la búsqueda de los vectores de la intersección del sistema de hiperplanos (46.9). Supongamos que los hiperplanos contienen por lo menos un vector común y convengamos en considerar, para simplificar, que el espacio lineal es real.

Elijamos un vector arbitrario v_0 y proyectémoslo sobre el primer hiperplano. El vector obtenido v_1 lo proyectemos sobre el segundo hiperplano, etc. Este proceso determina cierta sucesión $\{v_p\}$. Investiguémosla.

El elemento principal de un proceso de cálculo consiste en proyectar cierto vector v_p sobre un hiperplano definido por la ecuación (46.8). Está claro que el vector v_{p+1} satisface esta ecuación y está ligado con el vector v_p mediante la igualdad

$$v_{p+1} = v_p + tn$$

para cierto número t. Al sustituir v_{p+1} en la ecuación (46.8), hallaremos t. De aquí obtenemos que

$$v_{p+1} = v_p + \left(\frac{b_r - (n, v_p)}{(n, n)}\right) n.$$

De esta fórmula se desprende que todos los vectores de la sucesión $\{v_p\}$ se disponen en un plano obtenido mediante el desplazamiento de una cápsula lineal $L(n_1, n_2, \ldots, n_l)$ al vector v_0. Pero todos los vectores pertenecientes a la intersección de los hiperplanos (46.9) se ubican en otro plano, obtenido por desplazamiento del complemento ortogonal $L^{\perp}(n_1, n_2, \ldots, n_l)$. Existe un único vector z_0 que pertenece a ambos planos.

Si demostramos que una subsucesión de $\{v_p\}$, cualquiera que sea, converge a cierto vector perteneciente a los hiperplanos (46.9), entonces, por ser el plano cerrado, dicha sucesión convergerá precisamente a z_0. En este caso al vector z_0 convergerá también toda la sucesión $\{v_p\}$.

Para cualquier r los vectores $z_0 - v_{r+1}$, $v_{r+1} - v_r$ son ortogonales, por lo cual, de acuerdo con el teorema de Pitágoras,

$$\rho^2(z_0, v_r) = \rho^2(z_0, v_{r+1}) + \rho^2(v_r, v_{r+1}).$$

Sumando las igualdades obtenidas respecto de r, desde 0 hasta $p-1$, hallamos

$$\rho^2(z_0, v_0) = \rho^2(z_0, v_p) + \sum_{r=0}^{p-1} \rho^2(v_r, v_{r+1}).$$

Por consiguiente,

$$\sum_{r=0}^{p-1} \rho^2(v_r, v_{r+1}) \leqslant \rho^2(z_0, v_0),$$

de donde concluimos que

$$\rho(v_p, v_{p+1}) \to 0. \qquad (55.1)$$

Denotemos mediante H_r el hiperplano en la r-ésima fila (46.9). Está claro que la distancia del vector v_p hasta H_r no es superior a la distancia entre v_p y cualquier vector de H_r. De acuerdo con la construcción de $\{v_p\}$, entre cualesquiera k vectores sucesivos suyos hay sin falta un vector perteneciente a cualquiera de los hiperplanos. Haciendo uso de la desigualdad triangular y correlación límite (55.1), obtenemos

$$\rho(v_p, H_r) \leqslant \rho(v_p, v_{p+1}) + \rho(v_{p+1}, v_{p+2}) + \dots + \rho(v_{p+k-1}, v_{p+k}) \to 0 \qquad (55.2)$$

para todo $r = 1, 2, \dots, k$.

La sucesión $\{v_p\}$ es, evidentemente, acotada. Elijamos de ella una subsucesión convergente. Supongamos que la subsucesión converge al vector z_0'. Pasando al límite en (55.2), encontramos que

$$\rho(z_0', H_r) = 0$$

para todo $r = 1, 2, \dots, k$. Pero, como ya se ha observado más arriba, el vector z_0' debe coincidir con z_0. Por consiguiente, la sucesión $\{v_p\}$ converge a z_0.

Ejercicios.

1. ¿Se han usado en su esencia los conceptos de completitud y de carácter cerrado en la investigación realizada?
2. ¿Cómo se pueden hallar otras soluciones del sistema, si existen?
3. ¿Cuál es el comportamiento del proceso, si el sistema es incompatible?

PARTE II OPERADORES LINEALES

CAPÍTULO 7 MATRICES Y OPERADORES LINEALES

§ 56. Operadores

El elemento principal en la creación de los fundamentos del análisis matemático consiste en la introducción del concepto de función. De acuerdo con la definición, para prefijar una función es preciso indicar dos conjuntos X, Y de números reales y señalar una regla, según la cual a todo número $x \in X$ se le pone en correspondencia el número único $y \in Y$. Esta regla representa precisamente una función unívoca de la variable real x, definida en el conjunto X.

Al realizar la idea general de la dependencia funcional, no es totalmente obligatorio exigir que X, Y sean unos conjuntos de números reales. Entendiendo por X, Y los más diversos conjuntos de elementos, llegamos a la siguiente definición que generaliza el concepto de función.

La regla, de acuerdo con la cual a todo elemento x de cierto conjunto no vacío X se le pone en correspondencia un único elemento y del conjunto no vacío Y, se llama *operador*. El resultado y de la aplicación del operador A al elemento x se designa así:

$$y = A(x), \qquad y = Ax \tag{56.1}$$

y se dice que el operador A *actúa de* X *en* Y o bien *aplica* X *en* Y.

El conjunto X se llama *campo de definición* del operador A. El elemento y de (56.1) se denomina *imagen* del elemento x y el propio x, *preimagen** del elemento y. La totalidad T_A de todas las imágenes se llama *campo de valores* (o *imagen*) del operador A. En el caso cuando cualquier elemento $y \in Y$ tiene preimagen y ésta es la única, la

* En algunas obras se emplea el término «imagen recíproca». (*N. del Tr.*)

regla (56.1) se llama *biunívoca*. El operador se denomina, además, *aplicación*, *transformación* u *operación*.

En lo que sigue consideraremos, en lo esencial, solamente los así llamados *operadores lineales*. Las peculiaridades distintivas de estos últimos consisten en lo siguiente. En primer lugar, el campo de definición de un operador lineal es siempre cierto espacio lineal o subespacio. En segundo lugar, las propiedades del operador lineal están íntimamente relacionadas con las operaciones sobre los vectores de un espacio lineal. Al estudiar los operadores lineales supondremos, generalmente, que los espacios se dan sobre un campo de números reales o complejos. Por operador se entenderá en lo sucesivo un operador lineal, siempre que no haya especificaciones especiales. En la teoría general de operadores los operadores lineales desempeñan un papel de la misma importancia que una línea recta y un plano desempeñan en el análisis matemático. A esto se debe, de hecho, la necesidad en una investigación detallada de ellos.

Sean dados los espacios lineales X, Y sobre un mismo campo P. Consideremos un operador A, cuyo campo de definición constituye el espacio X y el campo de valores es cierto conjunto de Y. El operador A se llama *lineal*, si

$$A(\alpha u + \beta v) = \alpha Au + \beta Av \qquad (56.2)$$

para cualesquiera vectores u, $v \in X$ y cualesquiera números α, $\beta \in P$.

Nos hemos encontrado, ya más de una vez, con los operadores lineales. De acuerdo con (9.8), como operador lineal sirve la magnitud de un segmento dirigido. Su campo de definición representa en él el conjunto de todos los segmentos dirigidos de un eje, el campo de valores coincide con el conjunto de todos los números reales. Según se deduce de (21.2), una correspondencia isomorfa entre dos espacios lineales también será un operador lineal. Fijemos un subespacio L en un espacio lineal provisto de un producto escalar. Obtendremos dos operadores lineales, si a todo vector del espacio le asignamos o bien su proyección sobre el subespacio L o bien la perpendicular trazada desde este vector a L. La validez de esta afirmación se desprende de (30.5), (30.6).

Un operador que a todo vector x del espacio X pone en correspondencia el vector nulo del espacio Y es, evidentemente, lineal. Se llama operador *nulo* y se denota mediante el símbolo 0. Así pues,

$$0 = 0x.$$

Pondremos en correspondencia a todo vector $x \in X$ el mismo vector x. Se obtendrá un operador lineal E que actúa de X en X. Este operador se denomina *idéntico* u operador *unidad*. Por definición,

$$x = Ex.$$

Sea un operador lineal A que actúa del espacio X en el espacio Y. Construyamos un operador nuevo B, conforme a la inscripción $Bx = -Ax$. El operador obtenido B es también un operador lineal que actúa de X en Y. Se llama operador *opuesto* al operador A.

Fijemos, por fin, un número arbitrario α y a todo vector $x \in X$ lo pondremos en correspondencia el vector $\alpha x \in X$. Un operador construido de este modo será, por supuesto, lineal. Se llama operador *escalar*. Cuando $\alpha = 0$ obtenemos un operador nulo, cuando $\alpha = 1$ obtenemos un operador idéntico.

A continuación expondremos el método general para obtener operadores lineales, pero ahora demos a conocer algunas de sus peculiaridades características. Según se deduce de (56.2), la correlación

$$A\left(\sum_{i=1}^{r}\alpha_i x_i\right)=\sum_{i=1}^{r}\alpha_i A x_i$$

tiene lugar para cualesquiera vectores x_i y los números α_i. De aquí se deduce, en particular, que todo operador lineal A transforma un vector nulo en vector nulo, es decir,

$$0 = A0.$$

El campo de valores T_A del operador lineal A es un *subespacio* del espacio Y. Si $z = Au$, $w = Av$, entonces el vector $\alpha z + \beta w$ será, evidentemente, la imagen del vector $\alpha u + \beta v$ para cualesquiera números α, β. Por consiguiente, el vector $\alpha z + \beta w$ pertenece al campo de valores del operador A. La dimensión del subespacio T_A se llama *rango* del operador y se indica con r_A.

A la par con T_A examinemos un conjunto N_A de vectores $x \in X$ que satisfacen la igualdad

$$Ax = 0.$$

Este conjunto es también un *subespacio* y se denomina *núcleo* del operador A. La dimensión n_A del núcleo se denomina *defecto* del operador A.

El rango y el defecto no son características independientes del operador lineal A. Sea un espacio X de dimensión m. Descompongámoslo en la suma directa

$$X = N_A \dotplus M_A, \qquad (56.3)$$

donde N_A es el núcleo del operador A y M_A, cualquier subespacio complementario. Tomemos un vector arbitrario $x \in X$ y representémoslo en forma de una suma

$$x = x_N + x_M,$$

donde $x_N \in N_A$, $x_M \in M_A$. Si $y = Ax$, en virtud de la linealidad del operador A y de la condición $Ax_N = 0$, obtenemos que

$$y = Ax_M.$$

Por consiguiente, todo vector de T_A tiene por lo menos una preimagen de M_A.

Esta preimagen en M_A es en realidad *única*. Supongamos que para un vector $y \in T_A$ tenemos dos preimágenes $x'_M, x''_M \in M_A$. Ya que M_A es un subespacio, entonces $x'_M - x''_M \in M_A$. Pero en vista de que x'_M y x''_M son preimágenes de un mismo vector y, se tiene $x'_M - x''_M \in N_A$. Sólo el vector nulo es común para los subespacios M_A y N_A. Por esta razón $x'_M - x''_M = 0$, es decir, $x'_M = x''_M$.

De este modo, el operador A establece *una correspondencia biunívoca* entre los vectores de los subespacios T_A y M_A. En virtud de la linealidad del operador esta correspondencia es un isomorfismo. Por esta razón, las dimensiones de T_A y M_A coinciden y son iguales a r_A. De la descomposición (56.3) se infiere que

$$r_A + n_A = m. \tag{56.4}$$

Hemos de notar que el operador lineal A establece una correspondencia isomorfa entre el subespacio T_A y cualquier subespacio M_A de X, el cual en la suma directa con el núcleo del operador constituye todo el espacio X. Por esto podemos considerar que todo operador lineal A genera toda una familia de otros operadores lineales. En primer lugar, se trata del operador nulo definido en el núcleo N_A, es decir, del operador que actúa de N_A en 0. En segundo lugar, esto es un conjunto de operadores lineales que actúan de los subespacios M_A, complementarios al núcleo, en el subespacio T_A. Una circunstancia de gran importancia es que cada uno de los nuevos operadores coincide en su campo de definición con el operador A. Si $N_A = 0$, entonces $M_A = X$ y todo el segundo conjunto de operadores coincide con el operador A. En cambio, si $N_A = X$, entonces A es un operador nulo. Estas cuestiones las analizaremos nuevamente más adelante.

Ejercicios.

Demuéstrese que los siguientes operadores son lineales.

1. En un espacio lineal X viene dada una base. El operador A asigna a todo vector $x \in X$ su coordenada de número fijado.

2. En un espacio X provisto de un producto escalar se fija un vector x_0. El operador A asigna a todo vector $x \in X$ el producto escalar (x, x_0).

3. En un espacio V_3 se fija un vector x_0. El operador A asigna a todo vector $x \in V_3$ el producto escalar $[x, x_0]$.

4. El espacio X está formado por unos polinomios de coeficientes reales. El operador A asigna a todo polinomio su k-ésima derivada. Este operador lleva el nombre de operador de *diferenciación k-múltiple*.

5. En el espacio de los polinomios dependientes de la variable t el operador A asigna a todo polinomio $P(t)$ un polinomio $t \cdot P(t)$.

6. El espacio X está descompuesto en la suma directa de los subespacios S y T. Representemos todo vector $x \in X$ en forma de una suma $x = u + v$, donde $u \in S$, $v \in T$. El operador A asigna al vector x el vector u. Este operador lleva el nombre de operador de *proyección* sobre el subespacio S paralelamente al subespacio T.

§ 57. Espacio lineal de los operadores

Fijemos dos espacios lineales X, Y sobre un mismo campo P y examinemos el conjunto ω_{XY} de todos los operadores lineales que actúan de X en Y. En el conjunto ω_{XY} se pueden introducir operaciones de adición de operadores y de multiplicación de un operador por los números de P, transformando de este modo ω_{XY} en un espacio lineal.

Dos operadores A, B, que actúan de X en Y, son *iguales*, si se cumple la igualdad

$$Ax = Bx$$

para todo vector $x \in X$. Es fácil comprobar que la razón de igualdad de los operadores es una razón de equivalencia. La igualdad de los operadores se designa de tal modo:

$$A = B.$$

El operador C se llama *suma* de los operadores A, B que actúan de X en Y, si se verifica la igualdad

$$Cx = Ax + Bx$$

para todo vector $x \in X$. La suma de los operadores se indica

$$C = A + B.$$

Por definición, se pueden sumar cualesquiera operadores que actúan de X en Y. Si A, B son unos operadores lineales de ω_{XY}, su suma será también un operador lineal de ω_{XY}. Para cualesquiera vectores u, $v \in X$ y cualesquiera números α, $\beta \in P$ se tiene

$$C(\alpha u + \beta v) = A(\alpha u + \beta v) + B(\alpha u + \beta v) =$$
$$= \alpha Au + \beta Av + \alpha Bu + \beta Bv =$$
$$= \alpha(Au + Bu) + \beta(Av + Bv) = \alpha Cu + \beta Cv.$$

La operación de adición de los operadores es *una operación algebraica*. Es, además, asociativa. En efecto, sean A, B, C tres operadores lineales arbitrarios de ω_{XY}. Entonces, para todo vector $x \in X$ se verifican las igualdades

$$((A + B) + C)x = (A + B)x + Cx = Ax + Bx + Cx =$$
$$= Ax + (Bx + Cx) = Ax + (B + C)x = (A + (B + C))x$$

Pero esto significa que

$$(A + B) + C = A + (B + C).$$

La operación de adición de los operadores es *conmutativa*. Si A, B son unos operadores cualesquiera de ω_{XY} y x es un vector

de X, entonces

$$(A+B)\,x = Ax + Bx = Bx + Ax = (B+A)\,x,$$

es decir,

$$A+B=B+A.$$

Ahora es fácil mostrar que el conjunto ω_{XY} con la operación de adición de los operadores introducida es un grupo abeliano. Este conjunto tiene por lo menos un elemento nulo, por ejemplo, el operador nulo. Todo elemento de ω_{XY} tiene por lo menos un elemento opuesto, por ejemplo, un operador inverso. Todo lo demás proviene del teorema 7.1.

Según se deduce del mismo teorema, la operación de adición de los operadores tiene su inversa. Llamémosla *sustracción* y hagamos uso de los símbolos y las propiedades aceptadas.

El operador C se llama *producto del operador* A, que actúa de X en Y, *por el número* λ del campo P, si se verifica la igualdad

$$Cx = \lambda\cdot Ax$$

para todo vector $x \in X$. Este producto se designa con

$$C = \lambda A.$$

El producto de un operador lineal de ω_{XY} por un número es también un operador lineal de ω_{XY}. En efecto, para cualesquiera vectores $u, v \in X$ y cualesquiera números $\alpha, \beta \in P$ tenemos

$$C(\alpha u + \beta v) = \lambda A(\alpha u + \beta v) = \lambda(\alpha Au + \beta Av) = \\ = \alpha(\lambda Au) + \beta(\lambda Av) = \alpha Cu + \beta Cv.$$

No es difícil convencerse de que en la operación de adición de los operadores y en la de multiplicación de un operador por un número se cumplen todas las propiedades que determinan un espacio lineal. Por consiguiente, el conjunto ω_{XY} de todos los operadores lineales, que actúan del espacio lineal X en el espacio lineal Y, forma un espacio lineal nuevo. De aquí se infiere que desde el punto de vista de las operaciones de multiplicación de un operador por un número, de adición y sustracción de los operadores, se cumplen todas las reglas para las transformaciones equivalentes de las expresiones algebraicas operacionales. En lo sucesivo dichas reglas ya no serán un objeto de especificación especial.

Hemos de notar que nunca usamos la relación mutua de los espacios lineales X, Y. Pueden ser tanto diferentes como coincidentes. El conjunto ω_{XX} de los operadores lineales, que actúan del espacio X en el mismo espacio X, será uno de los fundamentales objetos de nuestra investigación. Los operadores citados se llamarán *operadores lineales en* X.

Ejercicios.

1. Demuéstrese que al multiplicar un operador por un número no nulo, el rango y el defecto del operador no varían.
2. Demuéstrese que el rango de la suma de unos operadores no es superior a la suma de rangos de los sumandos.
3. Demuéstrese que un conjunto de operadores lineales de ω_{XY}, cuyos campos de valores pertenecen a un mismo subespacio, forma de por sí un subespacio lineal.
4. Demuéstrese que un sistema de dos operadores no nulos de ω_{XY}, cuyos campos de valores son distintos, es linealmente independiente.
5. Demuéstrese que un espacio de operadores lineales que actúan en V_1, es unidimensional.

§ 58. Anillo de los operadores

Consideremos tres espacios lineales X, Y, Z sobre un mismo campo P. Sea A un operador que actúa de X en Y y sea B un operador que actúa de Y en Z.

El operador C que actúa de X en Z se llama *producto del operador B por el operador A*, si se verifica la igualdad

$$Cx = B(Ax)$$

para todo vector $x \in X$. El producto de los operadores B y A se designa por

$$C = BA.$$

El producto de los operadores lineales es también un operador lineal. Para cualesquiera vectores $u, v \in X$ y cualesquiera números $\alpha, \beta \in P$ se tiene

$$C(\alpha u + \beta v) = B(A(\alpha u + \beta v)) = B(\alpha Au + \beta Av) =$$
$$= \alpha B(Au) + \beta B(Av) = \alpha Cu + \beta Cv.$$

La multiplicación de los operadores no es una operación algebraica, dado que el producto no está definido para todo par de operadores. No obstante, siendo factible, la operación de multiplicación de los operadores posee unas propiedades bien naturales. A saber:

$$\begin{aligned} &1)\ (AB)\,C = A\,(BC), \\ &2)\ \lambda\,(AB) = (\lambda B)\,A = B\,(\lambda A), \\ &3)\ (A+B)\,C = AC + BC, \\ &4)\ A\,(B+C) = AB + AC \end{aligned} \qquad (58.1)$$

para cualesquiera operadores A, B, C y todo número λ de P, si, por supuesto, las expresiones correspondientes están definidas.

La demostración de todas estas propiedades se efectúa de una manera igual, razón por la cual nos limitaremos a estudiar solamente la primera de las propiedades. Sean X, Y, Z, U unos espacios

lineales fijados y sean A, B, C cualesquiera operadores lineales, de los cuales A actúa de X en Y, B actúa de Y en Z y C, de Z en U. Observemos ante todo que en la igualdad 1 están definidos ambos operadores, $(AB)\,C$ y $A\,(BC)$. Para todo vector $x \in X$ tenemos

$$((AB)\,C)\,x = AB\,(Cx) = A\,(B\,(Cx)),$$
$$(A\,(BC))\,x = A\,(BCx) = A\,(B\,(Cx)),$$

de donde se deduce la validez de la igualdad 1.

Consideremos otra vez el conjunto ω_{XX} de los operadores lineales que actúan en el espacio X. Para cualesquiera dos operadores de ω_{XX} se han definido tanto la suma como el producto. De acuerdo con las propiedades 3, 4, ambas operaciones están relacionadas por una ley distributiva. Por esto el conjunto ω_{XX} de operadores lineales representa en sí *un anillo*. En lo sucesivo mostraremos que el anillo de los operadores es *no conmutativo*. Desde luego, por casualidad puede ocurrir que para algún par de operadores A, B se cumpla la correlación $AB = BA$. Estos operadores se llamarán *conmutables*. En particular, el operador idéntico es conmutable con cualquier operador.

En el anillo de los operadores lineales, al igual que en todo otro anillo, el producto de cualquier operador por un operador nulo es también un operador nulo. La ley distributiva relaciona con la multiplicación no sólo la suma de operadores, sino también la diferencia entre éstos. Un anillo de los operadores lineales es a la vez un espacio lineal, por lo cual, para la diferencia entre los operadores resulta válida la fórmula

$$A - B = A + (-1)\,B.$$

La propiedad 2 de (58.1) muestra la relación entre la multiplicación de los operadores en un anillo y la multiplicación por un número. Desde luego, quedan en vigor también todas las correlaciones que provienen de las propiedades de los espacios lineales.

Ejercicios.

1. Denotemos mediante D el operador de diferenciación y mediante T, el operador de multiplicación por t, en un espacio de polinomios que dependen de la variable t. Demuéstrese que $DT \neq TD$. Hállese el operador $DT - TD$.

2. Fijemos cierto operador B del espacio ω_{XX}. Demuéstrese que el conjunto de los operadores A, para los cuales $BA = 0$, forma en ω_{XX} un subespacio.

3. Demuéstrese que el rango de un producto de los operadores no es superior al rango de cada uno de los factores.

4. Demuéstrese que el defecto de un producto de operadores no es inferior al defecto de cada uno de los factores.

5. Demuéstrese que en el anillo ω_{XX} de operadores lineales hay divisores de cero.

§ 59. Grupo de operadores regulares

Los operadores lineales que actúan en el espacio X forman un grupo abeliano de adición. Pero entre tales operadores pueden indicarse unos conjuntos que representan en sí los grupos de multiplicación. Estos grupos están ligados con los llamados operadores regulares.

Un operador que actúa en un espacio lineal se llama *regular*, si su núcleo consta sólo de un vector nulo. Un operador no regular se llama *degenerado*. Serán regulares, por ejemplo, el operador idéntico y el operador escalar, si no es nulo. A veces, con el operador A, que actúa en el espacio X, se puede ligar cierto operador regular, incluso cuando A sea degenerado. En efecto, sea T_A el campo de valores del operador A y sea N_A su núcleo. Si T_A y N_A no tienen vectores no nulos comunes, entonces, de acuerdo con (56.4), tenemos

$$X = N_A \dotplus T_A.$$

Como ya se ha observado, el operador A engendra un conjunto de otros operadores que actúan de cualquier subespacio, complementario al núcleo N_A, en el subespacio de valores T_A. En el caso considerado el operador A engendra un operador que actúa de T_A en T_A. Este operador será regular, puesto que transforma en cero sólo el vector nulo de T_A.

Los operadores regulares poseen una serie de peculiaridades remarcables. Para los operadores de este tipo el defecto es igual a cero, por lo cual de la fórmula (56.4) se infiere que el rango del operador regular coincide con la dimensión del espacio. Si un operador regular A actúa en el espacio X, el campo de valores T_A coincide con X. De este modo, todo vector de X es una imagen de cierto vector de X. Esta propiedad del operador regular es *equivalente* a su definición.

Una propiedad importante del operador regular consiste en la unicidad de la preimagen de todo vector del espacio. Efectivamente, supongamos que para cierto vector y existen dos preimágenes u, v. Esto significa que

$$Au = y, \qquad Av = y.$$

Pero en este caso

$$A(u - v) = 0.$$

Por definición de operador regular, el núcleo se compone sólo del vector nulo. Por esto, $u - v = 0$, esto es, $u = v$. La propiedad demostrada es también *equivalente* a la definición de operador regular. Esta propiedad ya se ha mencionado, de hecho, en el § 56.

Un producto de cualquier número finito de operadores regulares es también un operador regular. Evidentemente, basta demostrar

esta afirmación para dos operadores. Sean A, B cualesquiera operadores regulares que actúan en el mismo espacio X. Consideremos la ecuación

$$BAx = 0. \qquad (59.1)$$

De acuerdo con la definición de la multiplicación de operadores, esta ecuación significa que

$$B(Ax) = 0.$$

El operador B es regular, por lo cual de la última ecuación se deduce que $Ax = 0$. Pero A es también un operador regular y de aquí proviene que $x = 0$. Así pues, la ecuación (59.1) se satisface solamente por el vector nulo, es decir, el operador BA es regular.

Una suma de operadores regulares ya no será obligatoriamente un operador regular. Si A es un operador regular, lo será también el operador $(-1)A$. Pero la suma de estos operadores es un operador nulo que es degenerado.

Examinemos el conjunto de operadores regulares que actúan en un mismo espacio lineal. En dicho conjunto la multiplicación de los operadores es una operación *algebraica* y, además, *asociativa*. Entre los operadores regulares figura también el operador idéntico E, que desempeña el papel de la unidad. Efectivamente, es fácil comprobar que para todo operador A que actúa en el espacio X se verifica

$$AE = EA = A.$$

Si probamos que para todo operador regular A existe un operador regular que, siendo multiplicado por A, da un operador idéntico, esto será el testimonio de que el conjunto de todos los operadores regulares forma un grupo de multiplicación.

Sea A un operador regular. Como se sabe, para todo vector $y \in X$ existe un vector, y sólo uno, $x \in X$, relacionado con y mediante la expresión

$$y = Ax. \qquad (59.2)$$

Por consiguiente, a todo vector $y \in X$ se puede poner en correspondencia el único vector $x \in X$, para el cual y es una imagen suya. La correspondencia construida es un cierto operador. Se llama operador *inverso* del operador A y se designa mediante el símbolo A^{-1}. Si se verifica la desigualdad (59.2), entonces

$$x = A^{-1}y. \qquad (59.3)$$

Demostremos que el operador inverso es lineal y regular.

El producto está definido para cualesquiera operadores, no sólo para los lineales. Por esto, de la definición de operador inverso se desprende que

$$A^{-1}A = AA^{-1} = E. \qquad (59.4)$$

Para demostrar estas igualdades, basta aplicar a ambos miembros de (59.2) el operador A^{-1} y a cada uno de los miembros de (59.3), el operador A.

Tomemos unos vectores cualesquiera $u, v \in X$ y cualesquiera números $\alpha, \beta \in P$ y consideremos un vector

$$z = A^{-1}(\alpha u + \beta v) - \alpha A^{-1}u - \beta A^{-1}v.$$

Apliquemos ahora a ambos miembros de la igualdad el operador A. Tomando en consideración la linealidad del operador A y las correlaciones (59.4), concluimos que $Az = 0$. Puesto que el operador A es regular, esto quiere decir $z = 0$. Por consiguiente,

$$A^{-1}(\alpha u + \beta v) = \alpha A^{-1}u + \beta A^{-1}v,$$

es decir, el operador A^{-1} es lineal.

Es fácil mostrar que el operador A^{-1} es regular. Para todo vector y del núcleo del operador A^{-1} tenemos

$$A^{-1}y = 0.$$

Apliquemos a ambos miembros de esta igualdad el operador A. Como A es un operador lineal, tenemos $A0 = 0$. Teniendo presentes las correlaciones (59.4), concluimos que $y = 0$. Así pues, el núcleo del operador A^{-1} se compone sólo de un vector nulo, es decir, A^{-1} es el operador regular.

De este modo, el conjunto de operadores regulares representa en sí *un grupo de multiplicación*. Un poco más adelante mostraremos que este grupo es no *conmutativo*.

Con ayuda de los operadores regulares pueden construirse también unos grupos conmutativos. Sea A un operador arbitrario que actúe en el espacio X. Para cualquier número positivo entero p hallemos el p-ésimo *grado* del operador A mediante la igualdad

$$A^p = \overbrace{A \cdot A \ldots A}^{p}, \qquad (59.5)$$

donde en el segundo miembro están contenidos p factores. En virtud de la asociatividad de la operación de multiplicación, el operador A^p se define unívocamente. Desde luego, este operador es lineal.

Para cualesquiera números positivos y enteros p, r de (59.5) se deduce que

$$A^pA^r = A^{p+r}. \qquad (59.6)$$

Si se considera, por definición, que

$$A^0 = E$$

para todo operador A, entonces la fórmula (59.6) tendrá lugar para cualesquiera números no negativos y enteros p, r.

Supongamos que A es un operador regular, entonces para todo r no negativo será regular también el operador A^r. Por consiguiente,

para él existe un operador inverso. De acuerdo con las fórmulas (7.2), (59.5), tenemos

$$(A^r)^{-1} = (A^{-1})^r = \overbrace{A^{-1}A^{-1}\ldots A^{-1}}^{r}. \qquad (59.7)$$

Consideraremos también, por definición, que

$$A^{-r} = (A^r)^{-1}.$$

Teniendo en cuenta las fórmulas (59.5), (59.7) y tomando en consideración que $AA^{-1} = A^{-1}A$, no es difícil demostrar la correlación

$$A^pA^{-r} = A^{-r}A^p$$

para cualesquiera p, r no negativos y enteros. Esto significa que la fórmula (59.6) es válida para cualesquiera números enteros p, r.

Tomemos ahora un operador regular A y formemos el conjunto ω_A de los operadores del tipo A^p para todos los p enteros. En este conjunto la multiplicación de los operadores es una operación *algebraica* y, como se deduce de (59.6), *conmutativa*. Todo operador A^p tiene su inverso, igual a A^{-p}. En el conjunto ω_A figura también el operador idéntico E. Por consiguiente, el conjunto ω_A representa en sí *un grupo conmutativo de multiplicación*. Este grupo se denomina *cíclico*, generado por el operador A.

Ejercicios.

1. Demuéstrese que si para dos operadores lineales A, B de ω_{XX} se verifica la correlación $AB = E$, entonces ambos operadores son regulares.

2. Demuéstrese que para que los operadores A, B, de ω_{XX} sean regulares, es necesario y suficiente que sean regulares los operadores AB y BA.

3. Demuéstrese que si el operador A es regular y el número $\alpha \neq 0$, entonces el operador αA es también regular y $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$.

4. Demuéstrese que $T_A \subseteq N_A$ cuando, y sólo cuando, $A^2 = 0$.

5. Demuéstrese que para cualquier operador A se cumplen las correlaciones

$$N_A \subseteq N_{A^2} \subseteq N_{A^3} \ldots, \quad T_A \supseteq T_{A^2} \supseteq T_{A^3} \ldots$$

6. Demuéstrese que el operador P es un operador de proyección cuando, y sólo cuando, $P^2 = P$. ¿Qué representan en sí los subespacios N_P y T_P?

7. Demuéstrese que si P es un operador de proyección, entonces $E - P$ es también un operador de proyección.

8. Demuéstrese que si el operador A satisface la igualdad $A^m = 0$ para algún número positivo entero m, entonces el operador $\alpha E - A$ será regular para cualquier número $\alpha \neq 0$.

9. Demuéstrese que el operador lineal A, para el cual $E + \alpha_1 A + \alpha_2 A^2 + \ldots + \alpha_n A^n = 0$, es regular.

10. Demuéstrese que si A es un operador regular, entonces o bien todos los operadores en el grupo cíclico ω_A son distintos o bien cierta potencia del operador A coincide con el operador idéntico.

§ 60. Matriz del operador

Demos a conocer un método general para construir el operador lineal que actúa del espacio m-dimensional X en el espacio n-dimensional Y. Supongamos que a los vectores de la base $e_1, e_2, \ldots, e_m$ del espacio X les están asignados unos vectores $f_1, f_2, \ldots, f_m$ del espacio Y. En este caso *existe* un operador lineal A *y es*, además, *único*, que actúa de X en Y y que transforma todo vector e_k en el vector correspondiente f_k.

Supongamos que el operador buscado A existe. Tomemos un vector arbitrario $x \in X$ y representémoslo en forma de un desarrollo

$$x = \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_m e_m.$$

Entonces

$$Ax = A\left(\sum_{k=1}^{m} \xi_k e_k\right) = \sum_{k=1}^{m} \xi_k A e_k = \sum_{k=1}^{m} \xi_k f_k.$$

El segundo miembro de las correlaciones se determina unívocamente por el vector x y las imágenes de la base. Por eso la igualdad obtenida demuestra la unicidad del operador A, si éste existe. Por otra parte, podemos definir el operador A precisamente mediante esta igualdad, es decir, poner

$$Ax = \sum_{k=1}^{m} \xi_k f_k.$$

El operador obtenido, como es fácil de comprobar, es un operador lineal que actúa de X en Y y transforma, a la vez, todo vector e_k en el vector correspondiente f_k. El campo de valores T_A del operador A coincide con la cápsula lineal del sistema de vectores $f_1, f_2, \ldots, f_m$.

Ahora podemos enunciar una deducción importante: *el operador lineal A que actúa del espacio X en el espacio Y está enteramente definido mediante la totalidad de imágenes*

$$Ae_1, Ae_2, \ldots, Ae_m$$

para cualquier base fijada

$$e_1, e_2, \ldots, e_m$$

del espacio X.

Fijemos en el espacio X la base $e_1, e_2, \ldots, e_m$ y en el espacio Y, la base $q_1, q_2, \ldots, q_n$. El vector e_1 se transforma por el operador A en cierto vector Ae_1 del espacio Y, el cual, como todo vector de este espacio, puede ser desarrollado por vectores básicos

$$Ae_1 = a_{11}q_1 + a_{21}q_2 + a_{n1}q_n.$$

Análogamente,

$$Ae_2 = a_{12}q_1 + a_{22}q_2 + \ldots a_{n2}q_n,$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$Ae_m = a_{1m}q_1 + a_{2m}q_2 + \ldots + a_{nm}q_n.$$

Los coeficientes a_{ij} de estas correlaciones determinan una matriz A_{eq} de n filas y m columnas

$$A_{eq} = \begin{pmatrix} a_{11}a_{12}\ldots a_{1m} \\ a_{21}a_{22}\ldots a_{2m} \\ \cdots\cdots\cdots \\ a_{n1}a_{n2}\ldots a_{nm} \end{pmatrix}$$

que se denomina *matriz del operador A en bases elegidas.*

Como columnas de la matriz del operador sirven las coordenadas de los vectores $Ae_1, Ae_2, \ldots, Ae_m$ respecto de la base $q_1, q_2, \ldots, q_n$. Con el fin de determinar el elemento a_{ij} de la matriz del operador A hace falta aplicar el operador al vector e_j y tomar la i-ésima coordenada en la imagen Ae_j. Si, para abreviar, designamos mediante $\{x\}_i$ la i-ésima coordenada del vector x, entonces $a_{ij} = \{Ae_j\}_i$. En lo sucesivo haremos uso del método descrito para determinar los elementos de la matriz del operador.

Consideremos un vector arbitrario $x \in X$ y su imagen $y = Ax$. Aclaremos de qué modo se expresan las coordenadas del vector y en términos de las coordenadas del vector x y los elementos de la matriz del operador. Sea

$$x = \sum_{j=1}^{m} \xi_j e_j, \qquad y = \sum_{i=1}^{n} \eta_i q_i. \tag{60.1}$$

Calculamos

$$Ax = A\Big(\sum_{j=1}^{m} \xi_j e_j\Big) = \sum_{j=1}^{m} \xi_j A e_j = \sum_{j=1}^{m} \xi_j \sum_{i=1}^{n} a_{ij} q_i = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} \xi_j a_{ij}\Big) q_i.$$

Al comparar el segundo miembro de estas igualdades con el desarrollo (60.1) para el vector y, concluimos que deben cumplirse las igualdades

$$\sum_{j=1}^{m} a_{ij}\xi_j = \eta_i$$

para $i = 1, 2, \ldots, n$, es decir

$$\begin{aligned} a_{11}\xi_1 + a_{12}\xi_2 + \ldots + a_{1m}\xi_m &= \eta_1, \\ a_{21}\xi_1 + a_{22}\xi_2 + \ldots + a_{2m}\xi_m &= \eta_2, \\ \cdots\cdots\cdots\cdots\cdots \\ a_{n1}\xi_1 + a_{n2}\xi_2 + \ldots + a_{nm}\xi_m &= \eta_n. \end{aligned} \tag{60.2}$$

De este modo, todo operador lineal genera, cuando están fijadas las bases en los espacios X, Y, las correlaciones (60.2) que relacionan entre sí las coordenadas de la imagen y las de la preimagen. Con el fin de determinar las coordenadas de la imagen según las coordenadas de la preimagen, basta calcular los primeros miembros de estas correlaciones. Para determinar las coordenadas de la preimagen según las coordenadas conocidas del vector y hemos de resolver el sistema de ecuaciones algebraicas lineales (60.2) respecto de las incógnitas $\xi_1, \xi_2, \ldots, \xi_m$. La matriz de este sistema coincide con la matriz del operador.

Las correlaciones (60.2) establecen una relación profunda entre los operadores lineales y sistemas de ecuaciones algebraicas lineales. En particular, de (60.2) proviene que el rango del operador coincide con el de la matriz del operador y la dimensión del núcleo coincide con el número de soluciones fundamentales del sistema homogéneo reducido. De este hecho se deduce trivialmente la fórmula (56.4) y una serie de otras fórmulas.

A la relación existente entre los sistemas de ecuaciones algebraicas y los operadores lineales volveremos con frecuencia. Pero primeramente demostremos que entre los operadores y las matrices que, de hecho, determinan los sistemas del tipo (60.2) existe una correspondencia biunívoca. Ya hemos mostrado que todo operador A determina cierta matriz A_{qe}, siendo fijadas las bases. Tomemos ahora una matriz arbitraria A_{ne} de dimensiones $n \times m$. Siendo fijadas las bases en los espacios X, Y, las correlaciones (60.2) a todo vector $x \in X$ le ponen en correspondencia cierto vector $y \in Y$. Es fácil comprobar que esta correspondencia es un operador lineal. Construyamos la matriz del operador dado en las mismas bases. Todas las coordenadas del vector e_j son nulas, a excepción de la j-ésima coordenada que es igual a uno. De (60.2) se deduce que las coordenadas del vector Ae_j coinciden con los elementos de la j-ésima columna de la matriz A_{qe} y, por ende, $(Ae_j)_i = a_{ij}$. Por lo tanto, la matriz del operador construido coincide con la inicial A_{ne}.

Así pues, toda matriz $n \times m$ es una matriz de cierto operador lineal que actúa del espacio m-dimensional X en el espacio n-dimensional Y, para las bases fijadas en dichos espacios. De este modo *se establece una correspondencia biunívoca, para unas bases fijadas cualesquiera, entre los operadores lineales y las matrices rectangulares.* En este caso, tanto los espacios lineales como las matrices se consideran, desde luego, sobre el mismo campo P.

He aquí algunos de los ejemplos. Sea 0 un operador nulo. Tenemos

$$(0e_j)_i = (0)_i = 0.$$

Por consiguiente, todos los elementos de la matriz del operador nulo son iguales a cero. Tal matriz se llama *nula* y se designa por el símbolo 0.

Tomemos ahora un operador idéntico E. Para este operador encontramos

$$(Ee_j)_i = (e_j)_i = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases}$$

Por esta razón la matriz del operador idéntico tiene la forma siguiente. Es una matriz cuadrada en cuya diagonal principal se disponen las unidades, mientras que los demás lugares son ocupados por ceros. La matriz de un operador idéntico se denomina matriz *unidad* y se designa con la letra E.

Nos encontraremos frecuentemente con un tipo más de matrices. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ unos números arbitrarios del campo P. Construyamos una matriz cuadrada Λ la que tiene dichos números dispuestos por la diagonal principal, mientras que en los restantes lugares se disponen ceros, es decir,

$$\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Las matrices de tal índole se llaman *diagonales*. Si todos los elementos diagonales son iguales entre sí, la matriz se denomina *escalar*. En particular, la matriz unidad es escalar. Llamamos también diagonales las matrices rectangulares, construidas de modo análogo. Si nos referimos a las correlaciones (60.2), estableceremos fácilmente cómo actúa el operador lineal con la matriz Λ. Este operador "estira" la i-ésima coordenada de cualquier vector λ_i veces para todo i.

Ejercicios.

1. En un espacio de polinomios de grado no superior a n se halla fijada la base $1, t, t^2, \ldots, t^n$. ¿Qué forma tiene en esta base la matriz del operador de diferenciación?

2. En el espacio X se halla dado el operador P de proyección sobre el subespacio S, paralelamente al subespacio T. Fijamos en X cualquier base formada como la reunión de los subespacios S y T. ¿Qué forma tienen en dicha base las matrices de los operadores P y $E - P$?

3. Supongamos que el operador lineal A actúa de X a Y. Designemos mediante M_A un subespacio en X, complementario al núcleo N_A, y mediante R_A un subespacio en Y complementario a T_A. ¿Cómo variará la matriz del operador A, si al elegir las bases en X, Y, empleamos las bases de unos o de todos los subespacios indicados?

§ 61. Operaciones sobre las matrices

Ya se ha mostrado que, siendo fijados las bases en los espacios, todo operador lineal se define unívocamente por medio de su matriz. Por esto, las operaciones con los

operadores examinadas más arriba conducen a las operaciones bien determinadas sobre las matrices. En los problemas que actualmente son de interés para nosotros, la elección de una base no juega ningún papel, razón por la cual los operadores y sus matrices se designan mediante las mismas letras, omitiendo todos los índices concernientes a las bases.

Supongamos que dos operadores iguales actúan del espacio m-dimensional X en el espacio n-dimensional Y. Puesto que los operadores iguales se ponen de manifiesto de una manera igual, cualquiera que sea la situación, tendrán una misma matriz. Esto nos ofrece un fundamento para enunciar la siguiente definición.

Las matrices A, B de dimensiones iguales $n \times m$ con los elementos a_{ij}, b_{ij} se llaman *iguales*, si

$$a_{ij} = b_{ij}$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. La igualdad de las matrices se indica del modo siguiente:

$$A = B.$$

Supongamos ahora que del espacio X en el espacio Y actúan dos operadores A, B. Consideraremos el operador $C = A + B$. Designemos los elementos de las matrices de estos operadores con c_{ij}, a_{ij}, b_{ij}, respectivamente. Con arreglo a lo dicho anteriormente, $c_{ij} = \{Ce_j\}_i$. Teniendo presentes la definición de la suma de operadores y las propiedades de las coordenadas de los vectores respecto a las operaciones sobre ellos, obtendremos

$$c_{ij} = \{Ce_j\}_i = \{(A + B)e_j\}_i = \{Ae_j + Be_j\}_i =$$
$$= \{Ae_j\}_i + \{Be_j\}_i = a_{ij} + b_{ij}.$$

Por esta razón:

Se llama *suma* de dos matrices A, B de dimensiones iguales $n \times m$ con los elementos a_{ij}, b_{ij} una matriz C de las mismas dimensiones con los elementos c_{ij}, si

$$c_{ij} = a_{ij} + b_{ij}$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. La suma de las matrices se designa con

$$C = A + B.$$

Se llama *diferencia* de dos matrices A, B de dimensiones iguales $n \times m$ con los elementos a_{ij}, b_{ij} una matriz C de las mismas dimensiones con los elementos c_{ij}, si

$$c_{ij} = a_{ij} - b_{ij}$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. La diferencia entre las matrices se indica con

$$C = A - B.$$

Consideremos un operador A que actúa de X en Y y el operador $C = \lambda A$ para cierto número λ. Si a_{ij}, c_{ij} son los elementos de las matrices de dichos operadores, entonces

$$c_{ij} = (Ce_j)_i = (\lambda A e_j)_i = \lambda (Ae_j)_i = \lambda a_{ij},$$

y llegamos a la siguiente definición:

Se llama *producto de la matriz A de dimensiones $n \times m$ con los elementos a_{ij} por el número λ* una matriz C de las mismas dimensiones con los elementos c_{ij}, si

$$c_{ij} = \lambda a_{ij},$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. El producto de una matriz por un número se designa así

$$C = \lambda A.$$

Sean dados un espacio m-dimensional X y un espacio n-dimensional Y sobre un mismo campo P. Según lo demostrado más arriba, siendo fijadas las bases en X, Y, entre el conjunto ω_{XY} de todos los operadores que actúan de X en Y y el conjunto de todas las matrices de dimensiones $n \times m$ con los elementos del campo P tiene lugar una correspondencia biunívoca. Puesto que las operaciones sobre las matrices se introducían en concordancia con las operaciones sobre los operadores, el conjunto de las matrices $n \times m$, al igual que el conjunto ω_{XY}, representa en sí un espacio lineal.

Es fácil mostrar una de las bases del espacio de matrices. Ésta será, por ejemplo, el sistema de matrices $A^{(kp)}$ para $k = 1, 2, \ldots, n$, $p = 1, 2, \ldots, m$, donde los elementos $a_{ij}^{(kp)}$ de la matriz $A^{(kp)}$ se definen por las siguientes igualdades:

$$a_{ij}^{(kp)} = \begin{cases} 1, & \text{si } i = k,\ j = p, \\ 0 & \text{en todos los demás casos.} \end{cases}$$

En el espacio ω_{XY} de base sirve un sistema de operadores con matrices $A^{(kp)}$. De aquí concluimos que *un espacio lineal de operadores que actúan de X en Y es un espacio de dimensión finita y su dimensión es igual al producto mn.*

Supongamos dados tres espacios lineales X, Y, Z, con la particularidad de que el operador A actúa de X en Y, y el B, de Y en Z. Sean las dimensiones de los espacios citados m, n, p, respectivamente. Convengamos en considerar que en X, Y, Z están fijadas las bases $e_1, \ldots, e_m$, $q_1, \ldots, q_n$, $r_1, \ldots, r_p$. El operador A tiene la matriz $n \times m$ con los elementos a_{ij} y

$$Ae_j = \sum_{i=1}^{n} a_{ij} q_i.$$

El operador B tiene la matriz $p \times n$ con los elementos b_{ij} y

$$Bq_s = \sum_{k=1}^{p} b_{ks} r_k.$$

Al investigar la matriz del operador $C = BA$, llegamos a la conclusión que debe tener dimensiones $p \times m$, y sus elementos son como siguen:

$$c_{ij} = (Ce_j)_i = (BAe_j)_i = [B(\sum_{s=1}^{n} a_{sj} q_s)]_i =$$

$$= [\sum_{s=1}^{n} a_{sj} Bq_s]_i = [\sum_{s=1}^{n} a_{sj} \sum_{k=1}^{p} b_{ks} r_k]_i =$$

$$= [\sum_{k=1}^{p} (\sum_{s=1}^{n} b_{ks} a_{sj}) r_k]_i = \sum_{s=1}^{n} b_{is} a_{sj}.$$

La fórmula obtenida nos dicta la siguiente definición.

El producto de la matriz B de dimensiones $p \times n$ con los elementos b_{ij} por la matriz A de dimensiones $n \times m$ con los elementos a_{ij} es la matriz C de dimensiones $p \times m$ con los elementos c_{ij}, si

$$c_{ij} = \sum_{s=1}^{n} b_{is} a_{sj} \qquad (61.1)$$

para $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, m$. El producto de las matrices se designa con

$$C = BA.$$

De este modo, el producto se ha definido sólo para aquellas matrices, en las cuales el número de columnas del factor izquierdo equivale al número de filas del factor derecho. El elemento de la matriz del producto que se dispone en la intersección de la i-ésima fila y la j-ésima columna es igual a la suma de productos de todos los elementos de la i-ésima fila del factor izquierdo por los elementos correspondientes de la j-ésima columna del factor derecho.

Recordemos una vez más que entre los operadores lineales y las matrices tiene lugar una correspondencia biunívoca. Las operaciones sobre las matrices se introducían conforme a las operaciones con los operadores. Por esta razón, la multiplicación de matrices está ligada, mediante la correlación (58.1), con la sumación y la multiplicación de la matriz por un número.

Ya se ha observado que el anillo de operadores y el grupo de todos los operadores regulares, que actúan en un espacio lineal, son no conmutativos. Para demostrar esta afirmación es suficiente, evidentemente, hallar dos matrices cuadradas A, B tales que sea $AB \neq BA$.

Tomemos, por ejemplo,

$$A=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \qquad B=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}.$$

Calculamos:

$$AB=\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}, \qquad BA=\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix},$$

y el carácter no conmutativo de la multiplicación queda demostrado.

La operación de multiplicación de las matrices permite anotar cómodamente las correlaciones del tipo (60.2). Designamos con x_e la matriz de dimensiones $m \times 1$, compuesta por las coordenadas del vector x, y mediante y_q, la matriz de dimensiones $n \times 1$, compuesta por las coordenadas del vector y. Entonces, las correlaciones (60.2) serán equivalentes a una igualdad matricial

$$A_{qe}x_e = y_q. \tag{61.2}$$

Se llama igualdad *coordenada*, correspondiente a la igualdad *operacional*

$$Ax = y.$$

Dicha igualdad liga, en forma matricial, las coordenadas de la preimagen con las de la imagen a través de la matriz del operador.

Resulta importante observar que desde el punto de vista de la notación, las igualdades coordenada y operacional parecen ser completamente análogas, siempre que, desde luego, se omitan los índices, mientras que *el símbolo Ax* se entienda como el producto de A por x. Puesto que las notaciones y las propiedades de las operaciones sobre matrices y operadores coinciden, cualquier transformación de una igualdad operacional lleva a la misma transformación de la igualdad coordenada. Por esto, formalmente no importa de qué correlaciones se trata: de las matriciales o de las operacionales.

En adelante, de hecho, no haremos distinciones entre las igualdades operacionales y coordenadas. Más aún, *todos los conceptos y hechos nuevos referentes a los operadores, los extenderemos sin reservas a las matrices.*

Ejercicios.

1. Demuéstrese que las operaciones sobre las matrices están vinculadas a la operación de transposición mediante las siguientes correlaciones:

$$(\alpha A)' = \alpha A', \qquad (A + B)' = A' + B',$$
$$(AB)' = B'A', \qquad (A')' = A.$$

2. Demuéstrese que todo operador lineal de rango r puede representarse en forma de una suma de r operadores lineales y no puede representarse en forma de una suma de un número menor de operadores de rango 1.

3. Demuéstrese que una matriz de dimensiones $n \times m$ es de rango 1, si, y sólo si, puede representarse en forma de un producto de dos matrices no nulas de dimensiones $n \times 1$ y $1 \times m$.

4. Supongamos que para las matrices fijadas A, B se cumple la igualdad

$$AC = BC$$

para cualquier matriz C. Demuéstrese que $A = B$.

5. Hállese la forma general de una matriz cuadrada que conmutable con una matriz diagonal dada.

6. Demuéstrese que para que una matriz sea escalar, es necesario y suficiente que sea conmutable con todas las matrices cuadradas.

7. La suma de elementos diagonales de una matriz A se denomina *traza* de la matriz A y se designa con tr A. Demuéstrese que

$$\operatorname{tr} A = \operatorname{tr} A', \qquad \operatorname{tr} (\alpha A) = \alpha \cdot \operatorname{tr} A,$$
$$\operatorname{tr} (A + B) = \operatorname{tr} A + \operatorname{tr} B, \qquad \operatorname{tr} (BA) = \operatorname{tr} (AB).$$

8. Demuéstrese que una matriz real A es nula, si, y sólo si, $\operatorname{tr} (AA') = 0$.

§ 62. Matrices y determinantes

Las matrices desempeñan un papel esencial en la investigación de los operadores lineales. Como medio auxiliar en las investigaciones se utiliza con frecuencia un determinante. Consideraremos ahora algunas cuestiones relacionadas con las matrices y los determinantes.

Supongamos que un operador regular A actúa en el espacio X. Su rango coincide con la dimensión de X. Según se deduce de las fórmulas (60.2), esto significa que el rango del sistema de columnas de la matriz del operador coincide con su número. Esto es posible cuando, y sólo cuando el determinante de la matriz sea distinto de cero. Así pues,

Un operador que actúa en un espacio lineal será regular cuando, y sólo cuando, el determinante de su matriz sea distinto de cero.

La propiedad obtenida del operador regular sirve de base para las definiciones siguientes.

Una matriz cuadrada se llama *regular*, si su determinante es distinto de cero y se llama *degenerada*, en el caso contrario.

Naturalmente, apoyándose en las propiedades correspondientes de los operadores regulares, se puede decir, por ejemplo, que un producto de matrices regulares es, nuevamente, una matriz regular; todas las matrices regulares forman un grupo de multiplicación; cada matriz regular engendra un grupo cíclico, etc. La conexión existente con los operadores regulares permite afirmar que cualquier matriz regular A tiene, y además, una sola matriz A^{-1} tal, que

$$A^{-1}A = AA^{-1} = E. \tag{62.1}$$

La matriz A^{-1} se llama *inversa* de la matriz A.

Empleando el concepto de determinante, podemos indicar la forma explícita de los elementos de la matriz inversa a través de los menores de la matriz A. Sirvan de base para la resolución de este problema las fórmulas (40.5)—(40.9). Teniendo presente la fórmula (61.1) para un elemento del producto de dos matrices, concluimos que

o las ecuaciones (62.1) les satisface la matriz

$$A^{-1}=\begin{pmatrix} \frac{A_{11}}{d} & \frac{A_{21}}{d} & \dots & \frac{A_{n1}}{d} \\ \frac{A_{12}}{d} & \frac{A_{22}}{d} & \dots & \frac{A_{n2}}{d} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1n}}{d} & \frac{A_{2n}}{d} & \dots & \frac{A_{nn}}{d} \end{pmatrix}.$$

Aquí d es el determinante de la matriz A; A_{ij}, el complemento algebraico de un elemento a_{ij}. Debido a la unicidad de la matriz inversa, ésta, puede tener sólo esta forma.

Introduzcamos unas designaciones abreviadas para los menores de una matriz arbitraria A. El menor, dispuesto en las filas i_1, $i_2, \dots, i_p$ y en las columnas $j_1, j_2, \dots, j_p$, se designará mediante

$$A\begin{pmatrix} i_1 & i_2, \dots, i_p \\ j_1 & j_2, \dots, j_p \end{pmatrix}.$$

Consideraremos, además, que la coincidencia de algunos índices en la fila superior (inferior) de la designación del menor significa que son coincidentes las filas (columnas) del mismo menor.

TEOREMA 12.1. (*fórmula Binet—Cauchy*). *Supongamos que una matriz cuadrada C de orden n es igual al producto de dos matrices rectangulares A y B, cuyas dimensiones son $n \times m$ y $m \times n$, respectivamente, con la particularidad de que $m \geq n$.* Entonces

$$C\begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix} = \sum_{1 \leq k_1 < k_2 < \dots < k_n \leq m} A\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \dots & k_n \\ 1 & 2 & \dots & n \end{pmatrix}. \quad (62.2)$$

DEMOSTRACIÓN. Denotemos con a_{ij}, b_{ij}, c_{ij} los elementos de las matrices A, B, C. Por definición del producto de las matrices, tenemos

$$c_{ij} = \sum_{s=1}^{m} a_{is} b_{sj}.$$

Sustituyendo, los elementos de la matriz C por sus expresiones y sirviéndose de la propiedad de linealidad del determinante en relación a los vectores columna, encontramos

$$\det\begin{pmatrix} c_{11} & \dots & c_{1n} \\ c_{n1} & \dots & c_{nn} \end{pmatrix} = \det\begin{pmatrix} \sum_{s_1=1}^{m} a_{1s_1} b_{s_1 1} & \sum_{s_2=1}^{m} a_{1s_2} b_{s_2 2} & \dots & \sum_{s_n=1}^{m} a_{1s_n} b_{s_n n} \\ \dots & \dots & \dots & \dots \\ \sum_{s_1=1}^{m} a_{ns_1} b_{s_1 1} & \sum_{s_2=1}^{m} a_{ns_2} b_{s_2 2} & \dots & \sum_{s_n=1}^{m} a_{ns_n} b_{s_n n} \end{pmatrix} =$$

$$= \sum_{s_1=1}^{m} \det \begin{pmatrix} a_{1s_1}b_{s_11} & \sum\limits_{s_2=1}^{m} a_{1s_2}b_{s_22} & \ldots & \sum\limits_{s_n=1}^{m} a_{1s_n}b_{s_nn} \\ \ldots & \ldots & \ldots & \ldots \\ a_{ns_1}b_{s_11} & \sum\limits_{s_2=1}^{m} a_{ns_2}b_{s_22} & \ldots & \sum\limits_{s_n=1}^{m} a_{ns_n}b_{s_nn} \end{pmatrix} =$$

$$= \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \det \begin{pmatrix} a_{1s_1}b_{s_11} & a_{1s_2}b_{s_22} & \ldots & \sum\limits_{s_n=1}^{m} a_{1s_n}b_{s_nn} \\ \ldots & \ldots & \ldots & \ldots \\ a_{ns_1}b_{s_11} & a_{ns_2}b_{s_22} & \ldots & \sum\limits_{s_n=1}^{m} a_{ns_n}b_{s_nn} \end{pmatrix} = \ldots =$$

$$= \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \ldots \sum_{s_n=1}^{m} \det \begin{pmatrix} a_{1s_1}b_{s_11} & a_{1s_2}b_{s_22} & \ldots & a_{1s_n}b_{s_nn} \\ \ldots & \ldots & \ldots & \ldots \\ a_{ns_1}b_{s_11} & a_{ns_2}b_{s_22} & \ldots & a_{ns_n}b_{s_nn} \end{pmatrix} =$$

$$= \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \ldots \sum_{s_n=1}^{m} A\begin{pmatrix} 1 & 2 & \ldots & n \\ s_1 & s_2 & \ldots & s_n \end{pmatrix} b_{s_11}b_{s_22} \ldots b_{s_nn}. \qquad (62.3)$$

Cada uno de los índices $s_1, s_2, \ldots, s_n$ no depende de los otros y puede tomar cualesquiera valores desde 1 hasta m, por lo que la expresión obtenida representa en sí la suma de m^n sumandos. En dicha suma serán nulos aquellos sumandos cuyos valores tengan aunque sea dos índices iguales, puesto que serán iguales a cero los menores correspondientes de la matriz A. Todos los demás sumandos pueden dividirse en grupos con $n!$ sumandos en cada uno de ellos, reuniendo en cada grupo todos los sumandos, cuyos valores de los índices forman una misma totalidad de números.

Designemos mediante $k_1, k_2, \ldots, k_n$ la disposición de los valores de los índices $s_1, s_2, \ldots, s_n$, ordenada en forma creciente. Sea

$$\varepsilon(s_1, s_2, \ldots, s_n) = (-1)^N,$$

donde N es el número de transposiciones necesarias para transformar la permutación $s_1, s_2, \ldots, s_n$ a la $k_1, k_2, \ldots, k_n$. En este caso, dentro de los límites de un grupo de los valores de los índices $s_1, s_2, \ldots, s_n$, la suma de los sumandos correspondientes de (62.3) será igual a

$$\sum \varepsilon(s_1, s_2, \ldots, s_n) A\begin{pmatrix} 1 & 2 & \ldots & n \\ k_1 & k_2 & \ldots & k_n \end{pmatrix} b_{s_11}b_{s_22} \ldots b_{s_nn} =$$

$$= A\begin{pmatrix} 1 & 2 & \ldots & n \\ k_1 & k_2 & \ldots & k_n \end{pmatrix} \sum \varepsilon(s_1, s_2, \ldots, s_n) b_{s_11}b_{s_22} \ldots b_{s_nn} =$$

$$= A\begin{pmatrix} 1 & 2 & \ldots & n \\ k_1 & k_2 & \ldots & k_n \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \ldots & k_n \\ 1 & 2 & \ldots & n \end{pmatrix}.$$

De la correlación obtenida deducimos la fórmula (62.2).

COROLARIO. *El determinante del producto de dos matrices cuadradas es igual al producto de los determinantes de los factores.*

La suma en la fórmula (62.2) constará, en el caso dado, de un sumando, por lo cual

$$C\begin{pmatrix}1 & 2 & \dots & n\\ 1 & 2 & \dots & n\end{pmatrix}=A\begin{pmatrix}1 & 2 & \dots & n\\ 1 & 2 & \dots & n\end{pmatrix}B\begin{pmatrix}1 & 2 & \dots & n\\ 1 & 2 & \dots & n\end{pmatrix}$$

o bien, lo que es igual,

$$\det C = \det A\cdot\det B.$$

COROLARIO. *Supongamos que una matriz cuadrada C de orden n es igual al producto de dos matrices rectangulares A y B cuyas dimensiones son $n\times m$ y $m\times n$, respectivamente, con la particularidad de que $m < n$. En tal caso, $\det C = 0$.*

En efecto, agreguemos a las matrices A y B $n-m$ últimas columnas nulas y, correspondientemente, $n-m$ filas a cada una de las matrices. Las matrices obtenidas se convierten en las cuadradas de orden n, mientras que sus determinantes serán nulos. El producto de estas matrices nos da la matriz C. Por ello, de acuerdo con el primer corolario, $\det C = 0$.

Ejercicios.

1. Demuéstrese que para cualquier matriz regular A se verifica la igualdad $(A^{-1})' = (A')^{-1}$.
2. Demuéstrese que para cualquier matriz regular A es válida la igualdad $\det(A^{-1}) = (\det A)^{-1}$.
3. Demuéstrese que para cualquier matriz cuadrada A de orden n se verifica la igualdad $\det(\alpha A) = \alpha^n \det A$.
4. Demuéstrese que si para las matrices cuadradas A, B se verifica la igualdad $AB = E$, entonces A es regular y $B = A^{-1}$.
5. Escríbase la fórmula del tipo (62.2) para un menor arbitrario del producto de dos matrices.
6. Demuéstrese que para cualquier matriz real A todos los menores principales de las matrices $A'A$ y AA' son no negativos.
7. Demuéstrese que el rango de un producto de matrices no es superior al rango de cada uno de los factores.
8. Demuéstrese que en una operación de multiplicación por una matriz regular el rango no varía.

§ 63. Paso a la otra base

Siendo fijadas las bases en los espacios, la igualdad coordenada permite investigar totalmente la acción de un operador lineal. Evidentemente, cuanto más simple sea la forma de la matriz de un operador, tanto más eficaz será la realización de dicha investigación. Generalmente las matrices de los operadores dependen de las bases y nuestra tarea inmediata consiste en aclarar esta dependencia.

Sean $e_1, e_2, \ldots, e_m$ y $f_1, f_2, \ldots, f_m$ dos bases de un mismo espacio m-dimensional X. Los vectores $f_1, f_2, \ldots, f_m$ se definen unívocamente mediante sus descomposiciones

$$
\begin{aligned}
f_1 &= p_{11}e_1 + p_{21}e_2 + \ldots + p_{m1}e_m,\\
f_2 &= p_{12}e_1 + p_{22}e_2 + \ldots + p_{m2}e_m,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
f_m &= p_{1m}e_1 + p_{2m}e_2 + \ldots + p_{mm}e_m
\end{aligned}
\tag{63.1}
$$

según los vectores $e_1, e_2, \ldots, e_m$. Los coeficientes p_{ij} determinan la matriz

$$
P = \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1m} \\ p_{21} & p_{22} & \ldots & p_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ p_{m1} & p_{m2} & \ldots & p_{mm} \end{pmatrix},
$$

la cual se llama *matriz de la transformación de coordenadas* al pasar de la base $e_1, e_2, \ldots, e_m$ a la base $f_1, f_2, \ldots, f_m$.

Tomemos un vector arbitrario $x \in X$ y descompongámoslo según los vectores de ambas bases. Sea

$$
x = \sum_{i=1}^{m} \xi_i e_i = \sum_{i=1}^{m} \eta_i f_i.
$$

De acuerdo con (63.1) tenemos

$$
\sum_{i=1}^{m} \xi_i e_i = \sum_{i=1}^{m} \eta_i f_i = \sum_{i=1}^{m} \eta_i \sum_{j=1}^{m} p_{ji} e_j =
$$
$$
= \sum_{j=1}^{m} \Big(\sum_{i=1}^{m} \eta_i p_{ji} \Big) e_j = \sum_{i=1}^{m} \Big(\sum_{j=1}^{m} \eta_j p_{ij} \Big) e_i.
$$

Comparando los coeficientes de e_i en el primero y segundo miembros de las correlaciones obtenidas, encontramos

$$
\xi_i = \sum_{j=1}^{m} p_{ij} \eta_j \tag{63.2}
$$

para $i = 1, 2, \ldots, m$. Estas fórmulas se denominan *fórmulas de transformación de las coordenadas*. Designemos, como hasta ahora, mediante x_e y x_f las matrices de dimensiones $m \times 1$, formadas por las coordenadas del vector x en las bases correspondientes. Las fórmulas (63.2) muestran que

$$
x_e = P x_f. \tag{63.3}
$$

La matriz de la transformación de coordenadas debe ser *regular*, puesto que en el caso contrario tendrá lugar la dependencia lineal entre sus columnas y, por tanto, entre los vectores $f_1, f_2, \ldots, f_m$. Por supuesto, cualquier matriz regular es una matriz de cierta trans-

formación de coordenadas definida mediante la igualdad (63.3). Al multiplicar a la izquierda la igualdad (63.3) por la matriz P^{-1}, obtendremos

$$x_f = P^{-1}x_e.$$

Supongamos ahora que en el espacio lineal X vienen dadas tres bases $e_1, e_2, \ldots, e_m$, $f_1, \ldots, f_m$ y $r_1, \ldots, r_m$. El paso de la primera base a la tercera puede realizarse con ayuda de dos procedimientos: o bien directamente de la primera a la tercera o bien primero de la primera a la segunda, y después de la segunda a la tercera. No es difícil establecer la conexión entre las matrices correspondientes de la transformación de coordenadas. De acuerdo con (63.3), tenemos:

$$x_e = Px_f, \qquad x_f = Rx_r, \qquad x_e = Sx_r.$$

De las primeras dos correlaciones se desprende

$$x_e = Px_f = P(Rx_r) = (PR)\,x_r,$$

de donde proviene que

$$S = PR.$$

De este modo, cuando las coordenadas se transforman de manera consecutiva, la matriz de la transformación resultante será igual al producto de matrices de las transformaciones intermedias.

Examinemos otra vez el operador lineal A que actúa de X en Y. Elijamos en el espacio X dos bases $e_1, \ldots, e_m$ y $f_1, \ldots, f_m$, y en el espacio Y otras dos bases $q_1, \ldots, q_n$ y $t_1, \ldots, t_n$. En las primeras dos bases a un mismo operador A le corresponde la igualdad coordenada

$$y_q = A_{qe}x_e, \tag{63.4}$$

y en las otras dos bases, la igualdad

$$y_t = A_{tf}x_f. \tag{63.5}$$

En concordancia con estos pares de bases, para un mismo operador A tenemos dos matrices A_{qe} y A_{tf}.

Designemos con P la matriz de la transformación de coordenadas al pasar de la base $e_1, \ldots, e_m$ a la base $f_1, \ldots, f_m$ y con Q, la matriz de la transformación de coordenadas al pasar de $q_1, \ldots, q_n$ a $t_1, \ldots$ $\ldots, t_n$. Se tiene

$$x_e = Px_f, \qquad y_q = Qy_t. \tag{63.6}$$

Sustituyendo estas expresiones para x_e, y_q en (63.4), obtenemos

$$Qy_t = A_{qe}Px_f,$$

de donde se deduce que

$$y_t = (Q^{-1}A_{qe}P)\,x_f.$$

Al comparar la igualdad obtenida con (63.5), concluimos que

$$A_{tf} = Q^{-1} A_{qe} P. \qquad (63.7)$$

Esto es precisamente la correlación buscada que liga las matrices de un mismo operador en diferentes bases.

Ejercicios.

1. Demuéstrese que al pasar a otras bases el rango de la matriz del operador no cambia.

2. Demuéstrese que el determinante de una matriz de un operador que actúa en un espacio lineal no depende de como se elige la base.

3. ¿Qué correspondencia puede establecerse entre unos operadores regulares, que actúan en el espacio X, y las transformaciones de coordenadas en el mismo espacio?

4. Llamemos homónimas dos bases de un espacio real, si el determinante de su matriz de la transformación de coordenadas es positivo. Demuéstrese que todas las bases pueden ser repartidas entre dos clases de bases homónimas.

5. Llamemos izquierda una clase de las bases homónimas y derecha, la otra. Compárense dichas clases con las descritas en el § 34.

§ 64. Matrices equivalentes y matrices semejantes

A todo operador lineal A que actúa del espacio X en el espacio Y le corresponde un conjunto de sus matrices que se define por la posibilidad de elegir diferentes bases en X e Y. La estructura de este conjunto puede ser sustancialmente diferente en dependencia de si son o no coincidentes X e Y.

Dos matrices rectangulares A y B de dimensiones iguales se denominan *equivalentes*, si existen dos matrices cuadradas regulares R y S tales que se verifique

$$B = RAS.$$

De (63.7) se deduce que dos matrices, correspondientes a un mismo operador lineal, siendo diferente la elección de las bases en X e Y, serán siempre equivalentes entre sí. No es difícil ver que la afirmación contraria es también cierta. A saber, dos matrices equivalentes corresponden siempre a un mismo operador lineal en las bases adecuadamente elegidas. De este modo, a todo operador lineal que aplica X en Y le corresponde una clase de matrices equivalentes.

TEOREMA 64.1. *Para que dos matrices rectangulares de dimensiones iguales sean equivalentes, es necesario y suficiente que tengan un mismo rango.*

DEMOSTRACIÓN. Al multiplicar una matriz cualquiera por otras matrices regulares, su rango no varía, razón por la cual las matrices equivalentes tienen rangos iguales. Supongamos ahora que dos matrices de dimensiones iguales tienen un mismo rango. Demostraremos que estas matrices son equivalentes. Demostraremos, además, que

cada matriz de rango r es equivalente a la matriz

$$I_r = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{pmatrix}.$$
$$\underbrace{\qquad\qquad}_{r}$$

Sea dada una matriz rectangular de dimensiones $n \times m$. Define cierto operador lineal A que aplica el espacio X de base $e_1, e_2, \ldots, e_m$ en el espacio Y de base $q_1, q_2, \ldots, q_n$. Designemos mediante r el número de vectores linealmente independientes entre las imágenes de los vectores de la base $Ae_1, Ae_2, \ldots, Ae_m$. Sin perturbar la generalidad, podemos considerar que los vectores $Ae_1, Ae_2, \ldots, Ae_r$ son linealmente independientes, puesto que se puede conseguirlo numerando adecuadamente los vectores de la base. Los restantes vectores, $Ae_{r+1}, \ldots, Ae_m$, se expresan linealmente en términos de los primeros

$$Ae_k = \sum_{j=1}^{r} \tau_{kj} Ae_j \tag{64.1}$$

para $k = r + 1, \ldots, m$. Definamos la base nueva $f_1, f_2, \ldots, f_m$ en X de la manera siguiente:

$$f_k = \begin{cases} e_k, & k = 1, 2, \ldots, r, \\ e_k - \sum\limits_{j=1}^{r} \tau_{kj} e_j, & k = r + 1, \ldots, m \end{cases} \tag{64.2}$$

En este caso, en virtud de (64.1), tenemos

$$Af_k = 0 \tag{64.3}$$

para $k = r + 1, \ldots, m$. Hagamos, luego,

$$Af_j = t_j \tag{64.4}$$

para $j = 1, 2, \ldots, r$. Los vectores $t_1, t_2, \ldots, t_r$ son, por hipótesis, linealmente independientes. Completémoslos con ciertos vectores $t_{r+1}, \ldots, t_n$ hasta obtener una base en Y y examinemos la matriz del operador A en las bases nuevas $f_1, \ldots, f_m$ y $t_1, \ldots, t_n$. Los coeficientes de la k-ésima columna de dicha matriz coinciden con las coordenadas del vector Af_k en la base $t_1, \ldots, t_n$. De conformidad con las correlaciones (64.3), (64.4), la matriz del operador A coincidirá con I_r.

La matriz inicial y la I_r corresponden a un mismo operador, por lo cual son equivalentes. Consecuentemente, todas las matrices de un mismo rango son equivalentes a la matriz I_r y por esta razón son equivalentes entre sí.

En el transcurso de la demostración del teorema hemos respondido a una pregunta muy importante: "*¿Cómo se deben elegir las bases en los espacios X e Y, para que la matriz del operador lineal tenga una forma más simple?*" Además, hemos mostrado la forma explícita de esta matriz más simple.

La respuesta, tan sencilla y eficaz, resultó ser posible debido a que las bases en X e Y podían escogerse *independientemente* una de la otra. Supongamos ahora que el operador A actúa en el espacio X. Por supuesto, podríamos considerar nuevamente las imágenes y preimágenes en diferentes bases, no obstante esto no parece ser natural aquí, puesto que tanto las imágenes como las preimágenes pertenecen a un mismo espacio. El empleo de diferentes bases complicaría considerablemente la investigación del modo con que actúa el operador sobre los vectores del espacio X. Si la base es una, las matrices P y Q en (63.6) coinciden. Por consiguiente, a todo operador lineal que actúa en un espacio lineal le corresponde una clase de matrices relacionadas por las correlaciones

$$B = P^{-1}AP \tag{64.5}$$

para diferentes matrices regulares P. Las matrices de tal índole se llaman *semejantes* y la matriz P se denomina *matriz de la transformación de semejanza*.

La cuestión referente a las condiciones bajo las cuales dos matrices pueden ser semejantes se resuelve con unas dificultades bastante grandes y la respuesta se obtendrá más adelante. También se ampliará la cuestión acerca de la forma de la matriz más simple entre todas las matrices semejantes. Los dos capítulos siguientes están dedicados precisamente a las investigaciones de estos problemas.

Ejercicios.

1. Demuéstrese que el criterio de equivalencia de las matrices y el criterio de semejanza son unas correlaciones de equivalencia.
2. Demuéstrese que las matrices semejantes poseen una traza y un determinante iguales.
3. Demuéstrese que en una misma transformación de semejanza un grupo cíclico de matrices regulares pasa en un grupo cíclico.
4. Demuéstrese que en una misma transformación de semejanza un subespacio lineal de matrices pasa en un subespacio lineal.
5. En un conjunto de matrices cuadradas del mismo orden examinemos un operador consistente en la transformación de semejanza de dichas matrices con una matriz fijada de la transformación de semejanza. Demuéstrese que este operador es lineal.
6. Demuéstrese que el conjunto de todos los operadores de una transformación de semejanza definidos sobre un mismo conjunto de matrices cuadradas del mismo orden forma un grupo de multiplicación.

CAPÍTULO 8 POLINOMIO CARACTERÍSTICO

§ 65. Valores propios y vectores propios

Supongamos que un operador lineal A actúa en el espacio X. Esto significa que a todo vector $x \in X$ se le pone en correspondencia un vector $y = Ax$ del mismo espacio X. Puede ocurrir que para cierto vector no nulo x la imagen y la preimagen sean colineales. Según veremos en lo sucesivo, una situación semejante permite simplificar considerablemente la investigación del operador.

El número λ se denomina *valor propio* y el vector no nulo x, *vector propio* del operador lineal A, siempre que estén ligados entre sí mediante la correlación $Ax = \lambda x$.

Hemos de notar que si x es un vector propio correspondiente al valor propio λ, todo vector colineal αx será también un vector propio para $\alpha \neq 0$. Si al valor propio λ le corresponden dos vectores propios x, y, entonces todo vector no nulo del tipo $\alpha x + \beta y$ también será un vector propio. Por definición, el vector nulo no es propio. Por esto el conjunto X_λ de *todos* los vectores propios que son combinaciones lineales de cualquier número de vectores propios dados, correspondientes a un mismo valor propio λ, no será un subespacio. En el caso de ampliar X_λ, al agregarle un vector nulo, X_λ se convertirá en un subespacio. Este último se denominará subespacio *propio* del operador A correspondiente al valor propio λ.

No es difícil comprender que todos los vectores no nulos del espacio X serán vectores propios de los operadores 0, E y αE. Cada uno de estos operadores tiene sólo un valor propio que es igual a 0, 1 y α, respectivamente, y, consecuentemente, por lo menos un subespacio propio que coincide con todo el espacio X. El operador proyector P cuenta con dos totalidades de vectores propios: todos los vectores pertenecientes al campo de valores del operador P y todos los vectores pertenecientes al campo de valores del operador $E - P$. A la primera totalidad de vectores propios le corresponde el valor propio $\lambda = 1$; a la segunda, el valor propio $\lambda = 0$. En efecto, como $P^2 = P$,

tenemos

$$P(Px) = P^2x = Px = 1\cdot Px,$$
$$P((E-P)x) = (P-P^2)x = (P-P)x = 0 = 0\cdot(E-P)x.$$

Por consiguiente, el operador proyector tiene al menos dos subespacios propios.

TEOREMA 65.1. *Un sistema de vectores propios $x_1, x_2, \ldots, x_m$ del operador A, que corresponden a los valores propios distintos dos a dos $\lambda_1, \lambda_2, \ldots, \lambda_m$, es linealmente independiente.*

DEMOSTRACIÓN. Los vectores propios son no nulos por definición, razón por la cual el teorema es justo, a ciencia cierta, para $m = 1$. Supongamos que es válido para cualquier sistema de $m-1$ vectores propios y no es válido para los vectores $x_1, x_2, \ldots, x_m$. En este caso el sistema de dichos vectores será linealmente independiente, es decir, para ciertos números $\alpha_1, \alpha_2, \ldots, \alpha_m$, no iguales a cero simultáneamente, se verifica la igualdad

$$\alpha_1x_1 + \alpha_2x_2 + \ldots + \alpha_mx_m = 0. \tag{65.1}$$

Supongamos que $\alpha_1 \neq 0$. Aplicando A a (65.1), obtendremos

$$\alpha_1\lambda_1x_1 + \alpha_2\lambda_2x_2 + \ldots + \alpha_m\lambda_mx_m = 0. \tag{65.2}$$

Al multiplicar (65.1) por λ_m y al restarla de (65.2), hallamos

$$\alpha_1(\lambda_1-\lambda_m)x_1 + \alpha_2(\lambda_2-\lambda_m)x_2 + \ldots$$
$$\ldots + \alpha_{m-1}(\lambda_{m-1}-\lambda_m)x_{m-1} = 0.$$

Conforme a la suposición inductiva, de aquí se deduce que todos los coeficientes de los vectores $x_1, x_2, \ldots, x_{m-1}$ son nulos. En particular, $\alpha_1(\lambda_1-\lambda_m) = 0$, lo que contradice la condición $\lambda_1 \neq \lambda_m$ y la suposición $\alpha_1 \neq 0$. Por consiguiente, el sistema de vectores $x_1, x_2, \ldots, x_m$ es linealmente independiente.

COROLARIO. *Todo operador lineal que actúa en un espacio m-dimensional no puede tener más de m valores propios distintos dos a dos.*

Es de mayor interés el caso en que el operador A en un espacio m-dimensional tiene m valores propios distintos dos a dos. De acuerdo con el teorema 65.1, en este caso podemos escoger una base del espacio compuesta íntegramente de los vectores propios del operador A.

El operador A que actúa en el espacio m-dimensional X se llama *operador de estructura simple*, si tiene m vectores propios linealmente independientes.

El hecho de que entre todos los operadores lineales destacamos los de estructura simple se explica de una manera sencilla. Estos operadores, y sólo ellos, tienen en cierta base las matrices *diagonales*. Efectivamente, sean $x_1, x_2, \ldots, x_m$ los vectores propios del operador A linealmente independientes. Al tomarlos en calidad de los vectores básicos del espacio X, construyamos en la base citada la matriz del

operador A. Tenemos

$$Ax_1 = \lambda_1 x_1,$$
$$Ax_2 = \lambda_2 x_2,$$
$$\dots\dots\dots$$
$$Ax_m = \lambda_m x_m.$$

Recordemos que los elementos de las columnas de la matriz del operador coinciden con las coordenadas de las imágenes de los vectores de la base. Por eso, la matriz A_λ del operador A tendrá en la base de los vectores propios la forma siguiente:

$$A_\lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_m \end{pmatrix}.$$

Ahora, si el operador A tiene en cierta base $x_1, x_2, \dots, x_m$ una matriz diagonal con ciertos números $\lambda_1, \lambda_2, \dots, \lambda_m$ (no *forzosamente distintos*) en la diagonal principal, entonces $x_1, x_2, \dots, x_m$ serán vectores propios del operador A correspondientes a los valores propios $\lambda_1, \lambda_2, \dots, \lambda_m$.

De este modo, los operadores de estructura simple, y sólo ellos, tienen en cierta base unas matrices diagonales. Esta base puede ser compuesta *solamente* de los vectores propios del operador A. La acción de cualquier operador de estructura simple se reduce siempre al "alargamiento" de las coordenadas del vector en la base dada. Si todos los operadores lineales tuvieran estructura simple, el problema referente a la elección de la base en la que la matriz del operador tenga una forma más sencilla sería resuelto por completo. No obstante, con los operadores de estructura simple no se agotan todos los operadores lineales.

Ejercicios.

1. Supongamos que el operador A tiene un vector propio x correspondiente al valor propio λ. Demuéstrese que para el operador

$$\alpha_0 E + \alpha_1 A + \dots + \alpha_n A^n,$$

donde $\alpha_0, \alpha_1, \dots, \alpha_n$ son ciertos números, el vector x será también propio, correspondiente, esta vez, al valor propio $\alpha_0 + \alpha_1\lambda + \dots + \alpha_n\lambda^n$.

2. Demuéstrese que los operadores A y $A - \alpha E$ tienen los mismos vectores propios, cualesquiera que sean el operador A y el número α.

3. Demuéstrese que el operador A es regular cuando, y sólo cuando, no tiene valores propios nulos.

4. Demuéstrese que los operadores A y A^{-1} tienen los mismos vectores propios, cualquiera que sea el operador regular A. ¿De qué modo están relacionados entre sí los valores propios de estos operadores?

5. Demuéstrese que si el operador A es de estructura simple, el operador

$$\alpha_0 E + \alpha_1 A + \ldots + \alpha_n A^n$$

es también de estructura simple.

6. Demuéstrese que el operador de diferenciación que actúa en un espacio de polinomios no es operador de estructura simple. Hállense los vectores propios y los valores propios de este operador.

7. Examinemos el operador de una transformación de semejanza con la matriz diagonal. Demuéstrese que dicho operador es de estructura simple. Hállense todos sus vectores propios y valores propios.

§ 66. Polinomio característico

No todo operador lineal tiene aunque sea un solo vector propio. Supongamos, por ejemplo, que un operador actúa en el espacio V_2 y realiza el giro de cada segmento dirigido alrededor del origen de coordenadas a 90° en el sentido contrahorario. Es evidente que en este caso la imagen y la preimagen nunca serán colineales y el operador no tendrá ni un solo vector propio. Para investigar la cuestión de existencia de los vectores propios, deduzcamos primero una ecuación la que satisfacen todos los valores propios del operador lineal.

Supongamos que el operador lineal A actúa en el espacio m-dimensional X prefijado sobre el campo P. Si el operador dispone de valor propio λ, correspondiente al vector propio x, entonces, por definición, queda cumplida la correlación $Ax = \lambda x$, o bien, lo que es igual,

$$(\lambda E - A)\,x = 0. \tag{66.1}$$

El vector x es no nulo, por lo cual de (66.1) se deduce que el operador $\lambda E - A$ es degenerado. De este modo, los valores propios del operador A son aquellos números λ de P, y sólo ellos, para los cuales el operador $\lambda E - A$ es degenerado.

Fijemos en el espacio X cierta base $e_1, e_2, \ldots, e_m$ y designemos mediante A_e la matriz del operador A en dicha base. El operador $\lambda E - A$ es degenerado cuando, y sólo cuando, sea degenerada su matriz $\lambda E - A_e$, es decir, cuando

$$\det(\lambda E - A_e) = 0. \tag{66.2}$$

La determinación de valores propios no ha sido ligada con la elección de la base en el espacio X. Por esto los números λ del campo P, que satisfacen la ecuación (66.2), tampoco deben depender de la base. En realidad, *no depende de la elección de la base* el primer miembro de (66.2), cualquiera que sea λ, aunque formalmente esta dependencia se ha observado. Supongamos que en cierta otra base $f_1, f_2, \ldots, f_m$ el operador A tiene la matriz A_f. Conforme a (64.5), las matrices A_e y A_f están ligadas entre sí mediante la correlación

$$A_f = Q^{-1} A_e Q$$

donde Q es una matriz regular. Ahora, con cualquier λ de P encontramos

$$\det(\lambda E - A_f) = \det(\lambda Q^{-1}EQ - Q^{-1}A_eQ) = \det(Q^{-1}(\lambda E - A_e)Q) =$$
$$= \det Q^{-1} \det(\lambda E - A_e) \det Q = (\det Q)^{-1} \det(\lambda E - A_e) \det Q =$$
$$= \det(\lambda E - A_e).$$

Al tomar en consideración la expresión del determinante de la matriz en términos de los elementos de ésta, es fácil entender que el primer miembro de (66.2) puede ser representado en la forma:

$$\det(\lambda E - A_e) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_{m-1}\lambda^{m-1} + \alpha_m\lambda^m. \qquad (66.3)$$

Los coeficientes $\alpha_0, \ldots, \alpha_m$ se calculan de tal o cual manera según los elementos de la matriz A_e y no dependen de λ. La potencia máxima de λ sólo figura en el producto de los elementos diagonales de la matriz $\lambda E - A_e$ y, por ello,

$$\alpha_m = 1.$$

De aquí la expresión explícita para dos coeficientes más. A saber,

$$\alpha_0 = (-1)^m \det A_e, \qquad \alpha_{m-1} = -\operatorname{tr} A_e.$$

Se puede suponer, en general, que al representar el determinante $\det(\lambda E - A_e)$ en polinomio de λ, empleando para ello diferentes métodos, obtendremos las expresiones de un tipo análogo al segundo miembro de (66.3), mas con distintos coeficientes α_i. Sin embargo, en lo que sigue se mostrará que la suposición citada no tiene lugar. Los coeficientes en el segundo miembro de (66.3) no dependen de cómo se realice su cálculo. Teniendo en cuenta que el determinante $\det(\lambda E - A_e)$ no depende de la base, llegamos a que todos los coeficientes $\alpha_0, \ldots, \alpha_{m-1}$ son, en realidad, las características del operador A. La función

$$f(\lambda) = \alpha_0 + \alpha_1\lambda + \ldots + \alpha_{m-1}\lambda^{m-1} + \lambda^m \qquad (66.4)$$

se denomina *polinomio característico del operador* A.

A todo operador lineal se le atribuye un polinomio característico. Lo recíproco es también cierto. Todo polinomio del tipo (66.4) es característico para cierto operador lineal. A título del último puede servir, por ejemplo, un operador cuya matriz A_e tiene en alguna base la forma siguiente:

$$A_e = \begin{pmatrix} -\alpha_{m-1} & -\alpha_{m-2} & \ldots & -\alpha_1 & -\alpha_0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \qquad (66.5)$$

Es fácil convencerse de esto por comprobación inmediata, recurriendo al teorema de Laplace para calcular el determinante det $(\lambda E - A_e)$. Las matrices del tipo (66.5) se llaman *matrices de Frobenius*.

Para que el número λ del campo P sea el valor propio del operador A, es necesario y suficiente que satisfaga la ecuación

$$\alpha_0 + \alpha_1\lambda + \ldots + \alpha_{n-1}\lambda^{n-1} + \lambda^n = 0,$$

es decir, que sea una raíz del polinomio característico. No en cualquier campo P ni mucho menos, todo polinomio con los coeficientes de P tiene aunque sea una sola raíz de P. Como ejemplo puede indicarse el polinomio $\lambda^2 + 1$, que no tiene raíces ni en el campo de números racionales ni en el campo de números reales.

El campo P se llama *algebraicamente cerrado*, si todo polinomio con los coeficientes de P tiene al menos una raíz de P.

Así pues, si un operador lineal actúa en un espacio dado sobre un campo algebraicamente cerrado, tiene sin falta por lo menos un solo vector propio. Pueden construirse los diversos ejemplos de los campos algebraicamente cerrados, no obstante, el mayor valor práctico lo tiene sólo uno de ellos, a saber, el campo de números complejos. Nuestras investigaciones más próximas están dedicadas a la demostración de que dicho campo es algebraicamente cerrado.

Ejercicios.

1. Hállese el polinomio característico para los operadores nulo e idéntico.
2. Hállese el polinomio característico para el operador de diferenciación.
3. ¿Será señal de igualdad de operadores la coincidencia de los polinomios característicos?
4. Demuéstrese que los operadores con las matrices A y A' tienen polinomios característicos iguales.
5. Supongamos que en cierta base el operador tiene la matriz (66.5). Hállense en la misma base las coordenadas de los vectores propios.
6. Demuéstrese que el operador con la matriz (66.5) tiene estructura simple cuando, y sólo cuando, el polinomio característico cuenta con n raíces distintas dos a dos.

§ 67. Anillo de polinomios

En algunos ejercicios y ejemplos ya hemos atraído la atención del lector a las propiedades algebraicas de los polinomios. En relación con el estudio del polinomio característico estas investigaciones serán continuadas.

Sea dado un campo arbitrario P. Consideraremos un conjunto de polinomios, es decir, de funciones del tipo

$$f(z) = a_0 + a_1 z + \ldots + a_n z^n, \qquad (67.1)$$

dependientes del argumento z, que toma los valores de P, y que tienen los coeficientes $a_0, \ldots, a_n$ de P. Convengamos en que $f(z)$ es un polinomio de *grado n*, siempre que $a_n \neq 0$ y los coeficientes de núme-

ros mayores sean nulos. El único polinomio que no tiene grado determinado es aquel cuyos coeficientes son todos iguales a cero. Llamémoslo polinomio *nulo* y lo designaremos con el símbolo 0.

Dos polinomios se considerarán *iguales*, si son iguales todos sus coeficientes de los argumentos de potencias iguales.

Sean dados, ahora, los polinomios $f(x)$ y $g(x)$ de grados n y s respectivamente. Denotemos

$$f(x) = a_0 + a_1 x + \ldots + a_{n-1}x^{n-1} + a_n x^n,$$
$$g(x) = b_0 + b_1 x + \ldots + b_{s-1}x^{s-1} + b_s x^s \qquad (67.2)$$

y supongamos, para concretar, que $n \geq s$. Se denomina *suma* $f(x) + g(x)$ de los polinomios $f(x)$ y $g(x)$ el polinomio

$$f(x) + g(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1} + c_n x^n,$$

donde $c_i = a_i + b_i$ para todo $i \leq s$, y $c_i = a_i$ para todo $i > s$. La potencia de la suma de los polinomios es igual a n, si $n > s$, pero será inferior a n, para $n = s$, siempre que $b_n = -a_n$.

Se llama *producto* $f(x) \cdot g(x)$ *de los polinomios* $f(x)$ *y* $g(x)$ el polinomio

$$f(x) \cdot g(x) = d_0 + d_1 x + \ldots + d_{n+s-1}x^{n+s-1} + d_{n+s}x^{n+s},$$

donde

$$d_i = \sum_{k+l=i} a_k b_l$$

para $i = 0, 1, \ldots, n + s$. El coeficiente d_i es la suma de productos de aquellos coeficientes de los polinomios $f(x)$ y $g(x)$ cuyos índices, siendo sumados, dan i. Por ejemplo,

$$d_0 = a_0 b_0, \qquad d_{n+s} = a_n b_s.$$

De la última igualdad se desprende que $d_{n+s} \neq 0$, por lo cual la potencia del producto de polinomios no nulos es igual a la suma de potencias de los factores. Por consiguiente, el producto de los polinomios no nulos es un polinomio no nulo.

Como caso particular del producto de los polinomios interviene *el producto* $\alpha f(x)$ *del polinomio* $f(x)$ *por el número* α, puesto que un número no nulo puede considerarse como un polinomio de grado nulo.

El conjunto de polinomios con las operaciones introducidas más arriba representa en sí un *anillo conmutativo*. No nos detendremos en la comprobación de la validez de todos los axiomas.

TEOREMA 67.1. *Para cualquier polinomio* $f(x)$ *y el polinomio no nulo* $g(x)$ pueden hallarse los únicos polinomios $q(x)$ y $r(x)$ tales que se verifica la ecuación

$$f(x) = g(x)\,q(x) + r(x), \qquad (67.3)$$

con la particularidad de que el grado de $r(x)$ *es inferior al de* $g(x)$ *o bien* $r(x)=0$.

DEMOSTRACIÓN. Supongamos que los polinomios $f(x)$ y $g(x)$ son de grado n y s. Si $n < s$ o bien $f(x)=0$, entonces en la descomposición (67.3) podemos hacer $q(x)=0$, $r(x)=f(x)$. Supongamos, por esto, que $n \geqslant s$.

Representemos los polinomios $f(x)$ y $g(x)$ conforme a (67.2) y hagamos

$$f(x)-\frac{a_n}{b_s}x^{n-s}g(x)=f_1(x). \tag{67.4}$$

Supongamos que el grado del polinomio $f_1(x)$ es igual a n_1, y su coeficiente mayor es $a_{n_1}^{(1)}$. Es obvio que $n_1 < n$. Si $n_1 \geqslant s$, hagamos

$$f_1(x)-\frac{a_{n_1}^{(1)}}{b_s}x^{n_1-s}g(x)=f_2(x). \tag{67.5}$$

Designaremos mediante n_2 el grado y mediante $a_{n_2}^{(2)}$ el coeficiente mayor del polinomio $f_2(x)$. Si $n_2 \geqslant s$, hagamos otra vez

$$f_2(x)-\frac{a_{n_2}^{(2)}}{b_s}x^{n_2-s}g(x)=f_3(x), \tag{67.6}$$

etc.

Los grados de los polinomios $f_1(x)$, $f_2(x)$, ... van decreciendo. Por ello, realizados un número finito de pasos, llegaremos a la igualdad

$$f_{k-1}(x)-\frac{a_{n_{k-1}}^{(k-1)}}{b_s}x^{n_{k-1}-s}g(x)=f_k(x), \tag{67.7}$$

en la cual el polinomio $f_k(x)$ o bien es nulo o bien su grado n_k es inferior a s. En este momento el proceso se para.

Sumando ahora todas las igualdades del tipo (67.4) – (67.7), obtendremos

$$f(x)-\left(\frac{a_n}{b_s}x^{n-s}+\frac{a_{n_1}^{(1)}}{b_s}x^{n_1-s}+\ldots+\frac{a_{n_{k-1}}^{(k-1)}}{b_s}x^{n_{k-1}-s}\right)g(x)=f_k(x).$$

Esto es testimonio de que los polinomios

$$q(x)=\frac{a_n}{b_s}x^{n-s}+\frac{a_{n_1}^{(1)}}{b_s}x^{n_1-s}+\ldots+\frac{a_{n_{k-1}}^{(k-1)}}{b_s}x^{n_{k-1}-s},\quad r(x)=f_k(x)$$

satisfacen la igualdad (67.3), con la particularidad de que o bien $r(x)=0$, o bien el grado del polinomio $r(x)$ es inferior al grado de $g(x)$.

Demostraremos ahora que los polinomios $q(x)$ y $r(x)$, que satisfacen la condición del teorema, son únicos. Supongamos que existen además los polinomios $q'(x)$ y $r'(x)$, para los cuales

$$f(x)=g(x)q'(x)+r'(x).$$

con la particularidad de que o bien $r'(z) = 0$, o bien el grado de $r'(z)$ es inferior al grado de $g(z)$. Entonces

$$g(z)\,(q(z) - q'(z)) = r'(z) - r(z). \qquad (67.8)$$

El polinomio en el segundo miembro de esta igualdad o bien es nulo o bien su grado es inferior al grado de $g(z)$. Mientras tanto, el polinomio en el primer miembro tiene, cuando $q(z) - q'(z) \neq 0$, un grado que no es inferior al grado de $g(z)$. Por esta razón, la igualdad (67.8) se verifica sólo en el caso en que

$$q(z) = q'(z), \qquad r(z) = r'(z).$$

El teorema queda demostrado por completo.

El polinomio $q(z)$ se denomina *cociente* de la división de $f(z)$ por $g(z)$, y $r(z)$ es el *resto* de la división. Si el resto es igual a cero, diremos que $f(z)$ se divide por $g(z)$ y el propio polinomio $g(z)$ se llamará *divisor* del polinomio $f(z)$.

Consideremos la división de un polinomio arbitrario no nulo $f(z)$ por un polinomio de primer grado $z - a$. Tenemos

$$f(z) = (z - a)\,q(z) + r(z). \qquad (67.9)$$

Puesto que el grado de $r(z)$ debe ser inferior al grado del polinomio $z - a$, entonces $r(z)$ es un polinomio de grado nulo, es decir, una constante. Esta constante se puede determinar con facilidad. Sustituyamos en el primero y segundo miembros de la correlación (67.9) $z = a$ y encontramos que $r(z) = f(a)$. Así,

$$f(z) = (z - a)\,q(z) + f(a). \qquad (67.10)$$

Para que el polinomio $f(z)$ se divida por el polinomio $z - a$, es necesario y suficiente que $f(a) = 0$. Los números a, para los cuales $f(a) = 0$, suelen llamarse *raíces* del polinomio $f(z)$. De este modo, la búsqueda de todos los divisores lineales del polinomio es equivalente a la búsqueda de todas sus raíces.

El empleo de la fórmula (67.10) permite hacer la siguiente deducción. Para todo número a de P un polinomio $f(z)$ de grado n puede ser representado de un modo único en forma de la descomposición en potencias $(z - a)$:

$$f(z) = A_0 + A_1(z - a) + \ldots + A_{n-1}(z - a)^{n-1} + A_n(z - a)^n, \qquad (67.11)$$

donde $A_0, \ldots, A_n$ son los números de P.

La existencia de aunque sea una sola descomposición (67.11) se establece de un modo relativamente sencillo. Al dividir $f(z)$ por $(z - a)$, obtendremos el cociente $q_1(z)$ y el resto A_0, relacionados mediante la igualdad

$$f(z) = (z - a)\,q_1(z) + A_0. \qquad (67.12)$$

Si el grado de $q_1(z)$ es nulo, la descomposición (67.11) queda obtenida. En cambio, si el grado de $q_1(z)$ es distinto de cero, entonces al dividir $q_1(z)$ por $(z-a)$, tendremos

$$q_1(z) = (z-a)\, q_2(z) + A_1. \tag{67.13}$$

Reuniendo (67.12), (67.13), encontramos

$$f(z) = (z-a)^2 q_2(z) + A_1 (z-a) + A_0.$$

Cuando sea necesario, dividimos otra vez $q_2(z)$ por $(z-a)$, etc. Como los grados de los cocientes $q_1(z)$, $q_2(z)$, ... decrecen de manera sucesiva, el proceso se parará dentro de n pasos, dando por resultado la descomposición (67.11).

Supongamos ahora que la descomposición del mismo tipo se ha obtenido de algún otro modo y tiene los coeficientes $A_0', \ldots, A_n'$. Al designar

$$q_i'(z) = A_i' + A_{i+1}'(z-a) + \ldots + A_n'(z-a)^{n-i}$$

para $i = 0, 1, \ldots, n$. Concluimos que

$$q_i'(z) = (z-a)\, q_{i+1}'(z) + A_i'. \tag{67.14}$$

En este caso, naturalmente, $q_0'(z) = f(z)$. Comparando (67.12) y (67.14) para $i = 0$, y tomando en consideración la unicidad del cociente y del resto, llegamos a la conclusión de que $A_0 = A_0'$, $q_1(z) = q_1'(z)$. De modo análogo se demuestra que los otros coeficientes son también iguales.

Ejercicios.

1. Demuéstrese que en el anillo de polinomios no hay divisores de cero.

2. Supongamos que para ciertos polinomios se verifica la igualdad $f(z)\varphi(z) = g(z)\varphi(z)$. Demuéstrese que si $\varphi(z) \neq 0$, entonces $f(z) = g(z)$.

3. Demuéstrese que los polinomios no nulos $f(z)$ y $g(z)$ se dividen uno por otro cuando, y sólo cuando, $g(z) = \alpha f(z)$ para el número α no nulo.

4. Supongamos que cada uno de los polinomios $f_1(z), \ldots, f_k(z)$ se divide por $\varphi(z)$. Demuéstrese que el polinomio $f_1(z) g_1(z) + \ldots + f_k(z) g_k(z)$, donde $g_1(z), \ldots, g_k(z)$ son unos polinomios arbitrarios, también se divide por $\varphi(z)$.

5. Demuéstrese que en las descomposiciones (67.1), (67.11) para un mismo polinomio $f(z)$ los coeficientes a_k y A_k coinciden.

§ 68. Teorema fundamental del álgebra

Procedamos a demostrar una de las más importantes afirmaciones, es decir, al teorema acerca de que el campo de números complejos es algebraicamente cerrado. Este teorema es de amplio uso en las más diversas ramas de las matemáticas. En particular, en este teorema está basada toda la teoría

ulterior de los operadores lineales. Conforme a la tradición establecida, se llamará *teorema fundamental del álgebra*.

Así pues, hemos de mostrar que todo polinomio de grado $n \geqslant 1$ con coeficientes complejos tiene al menos una raíz que es, en el caso general, compleja. Consideraremos en principio unos polinomios del tipo especial. A saber,

$$f(z) = a - z^n. \qquad (68.1)$$

Representemos los números complejos z en la así llamada *forma triangular*

$$z = r(\cos\varphi + i \operatorname{sen}\varphi).$$

Aquí, r es un número no negativo, llamado *módulo* del número z, y φ es un número real llamado *argumento* del número z. Está claro que para todo número z el módulo está definido unívocamente. Para los números z no nulos el módulo está definido, salvo un número múltiple de 2π; para $z = 0$ el argumento no está definido. Formando el producto de dos números complejos

$$z = r(\cos\varphi + i \operatorname{sen}\varphi), \qquad v = \rho(\cos\psi + i \operatorname{sen}\psi),$$

encontramos

$$zv = r\rho(\cos\varphi + i \operatorname{sen}\varphi)(\cos\psi + i \operatorname{sen}\psi) = r\rho(\cos(\varphi+\psi) + i \operatorname{sen}(\varphi+\psi)).$$

De aquí deducimos que

$$z^n = r^n(\cos n\varphi + i \operatorname{sen} n\varphi).$$

Esta igualdad lleva el nombre *de Moivre*. Permite hallar con facilidad las raíces de la ecuación (68.1). En efecto, supongamos que el número complejo a está representado en la forma triangular

$$a = \alpha(\cos\theta + i \operatorname{sen}\theta).$$

La ecuación

$$a - z^n = 0$$

respecto de z es equivalente a la ecuación

$$\alpha(\cos\theta + i \operatorname{sen}\theta) = r^n(\cos n\varphi + i \operatorname{sen} n\varphi)$$

respecto de r y φ. Mas, la última ecuación tiene, a ciencia cierta, tales soluciones para $k = 0, 1, 2, \ldots, n-1$:

$$r = +\sqrt[n]{\alpha}, \qquad \varphi = \frac{\theta + 2k\pi}{n}.$$

Por consiguiente, los números complejos

$$a_k = +\sqrt[n]{\alpha}\left(\cos\frac{\theta+2k\pi}{n} + i \operatorname{sen}\frac{\theta+2k\pi}{n}\right) \qquad (68.2)$$

son las raíces de la ecuación (68.1). Llamaremos estos números raíces de *n-ésimo grado del número* a y las designaremos mediante el sím-

bolo general

$$\alpha_k = \sqrt[n]{a}.$$

Ahora, sea dado un polinomio arbitrario $f(z)$ con coeficientes complejos. Considerarémoslo como una función compleja del argumento complejo z. Para tales funciones, al igual que para las funciones reales de un argumento real, pueden introducirse las nociones de continuidad, derivada, etc. De éstas no todas las nociones las necesitaremos en una medida igual, pero todas ellas se basan en el empleo de la complitud del espacio de números complejos.

Una función compleja uniforme $f(z)$ del argumento complejo z se llama *continua en el punto* z_0, si para todo número $\varepsilon > 0$, tan pequeño como se quiera, existe un $\delta > 0$ tal que para cualquier número complejo z que satisface la desigualdad

$$|z - z_0| < \delta$$

tendremos

$$|f(z) - f(z_0)| < \varepsilon.$$

La función $f(z)$, continua en todo punto del campo de definición, se denomina *continua en todo punto* o, simplemente, *continua*.

LEMA 68.1. *El polinomio $f(z)$ con coeficientes complejos es una función continua del argumento complejo z.*

DEMOSTRACION. Sea

$$f(z) = a_0 + a_1 z + \ldots + a_n z^n \tag{68.3}$$

y supongamos que z_0 es un número complejo arbitrario fijado. Designemos $h = z - z_0$. Mostremos que para cualquier número $\varepsilon > 0$, tan pequeño como se quiera, se puede hallar tal $\delta > 0$, que, siendo $|h| < \delta$, se cumpla la desigualdad $|f(z) - f(z_0)| < \varepsilon$.

Al desarrollar el polinomio dado $f(z)$ en potencias de $(z - z_0)$, obtendremos

$$f(z) = A_0 + A_1 (z - z_0) + \ldots + A_n (z - z_0)^n.$$

Puesto que $A_0 = f(z_0)$ y $(z - z_0)$ está designado con h, resulta

$$f(z_0 + h) - f(z_0) = A_1 h + \ldots + A_n h^n. \tag{68.4}$$

De aquí proviene que

$$|f(z_0 + h) - f(z_0)| \leqslant |A_1|\,|h| + \ldots$$
$$\ldots + |A_n|\,|h|^n = A(|h|). \tag{68.5}$$

La función real $A(|h|)$ es un polinomio con coeficientes reales $|A_i|$ respecto de la variable real $|h|$. Según se sabe del curso del análisis matemático, $A(|h|)$ es una función continua en todo punto y, en particular, cuando $|h| = 0$. Como $A(0) = 0$, por $\varepsilon > 0$

dado se puede hallar tal $\delta > 0$ que para

$$|h| < \delta \tag{68.6}$$

tendremos

$$A(|h|) < \varepsilon.$$

Tomando en consideración la desigualdad (68.5), concluimos que si se cumple (68.6), se cumplirá también la desigualdad

$$|f(z_0 + h) - f(z_0)| < \varepsilon.$$

COROLARIO. *El módulo de un polinomio es una función continua.*

Dicha afirmación se deduce inmediatamente de la siguiente correlación:

$$||f(z)| - |f(z_0)|| \leqslant |f(z) - f(z_0)|.$$

COROLARIO. *Si una sucesión de números complejos $\{z_k\}$ converge a z_0, para todo polinomio $f(z)$*

$$\lim_{k \to \infty} f(z_k) = f(z_0).$$

LEMA 68.2. *Si el polinomio $f(z)$ de grado $n \geqslant 1$ no se anula cuando $z = z_0$, entonces siempre existe un número complejo h tal, que*

$$|f(z_0 + h)| < |f(z_0)|.$$

DEMOSTRACIÓN. Examinemos nuevamente el desarrollo (68.4). Supongamos que entre los coeficientes $A_1, A_2, \ldots, A_n$ el primer coeficiente distinto de cero es A_k. Tomemos

$$h = t\sqrt[k]{-\frac{f(z_0)}{A_k}}, \tag{68.7}$$

donde a título de raíz de la k-ésima potencia se tome cualquiera de los valores de este coeficiente y

$$0 \leqslant t \leqslant 1. \tag{68.8}$$

Designaremos

$$B_p = A_p\left(\sqrt[k]{-\frac{f(z_0)}{A_k}}\right)^p.$$

Ahora, teniendo en cuenta (68.7), (68.8), de (68.4) encontramos

$$\begin{aligned} |f(z_0 + h)| &= |f(z_0) - t^k f(z_0) + t^{k+1}B_{k+1} + \ldots + t^n B_n| \leqslant \\ &\leqslant |(1 - t^k) f(z_0)| + t^{k+1}|B_{k+1}| + \ldots + t^n|B_n| = \\ &= (1 - t^k)|f(z_0)| + t^{k+1}|B_{k+1}| + \ldots + t^n|B_n| = \\ &= |f(z_0)| + t^k[-|f(z_0)| + t|B_{k+1}| + \ldots \\ &\quad \ldots + t^{n-k}|B_n|] = |f(z_0)| + t^k B(t). \end{aligned}$$

En definitiva tenemos

$$|f(z_0 + h)| \leqslant |f(z_0)| + t^k B(t).$$

La función $B(t)$ es un polinomio con coeficientes reales y el argumento real t. Es una función continua. Pero, $B(0) = -|f(z_0)| < 0$, razón por la cual, en virtud de la continuidad de $B(t)$, existe tal t_0 dentro de los límites $0 < t_0 \leq 1$, que $B(t_0)$ será también negativo. Para el número complejo h, que se determina mediante el número t_0, de acuerdo con (68.7) obtendremos

$$|f(z_0 + h)| \leq |f(z_0)| + t_0^k B(t_0) < |f(z_0)|.$$

LEMA 68.3. *Para todo polinomio $f(z)$ de grado $n \geq 1$ y toda sucesión infinita de números complejos $\{z_k\}$ se verifica la correlación límite*

$$\lim_{k \to \infty} |f(z_k)| = +\infty. \qquad (68.9)$$

DEMOSTRACIÓN. Consideraremos el polinomio (68.3). Para cualquier $z \neq 0$ encontraremos

$$|f(z)| \geq |a_n||z|^n \left(1 - \left|\frac{a_0}{a_n}\right||z|^{-n} - \ldots - \left|\frac{a_{n-1}}{a_n}\right||z|^{-1}\right). \qquad (68.10)$$

Puesto que la sucesión $\{z_k\}$ es infinitamente creciente, entonces

$$\lim_{k \to \infty} |z_k| = +\infty.$$

El segundo miembro de la correlación (68.10) es una función real, por lo cual calculamos

$$\lim_{|z_k| \to \infty} \left(1 - \left|\frac{a_0}{a_n}\right||z_k|^{-n} - \ldots - \left|\frac{a_{n-1}}{a_n}\right||z_k|^{-1}\right) = 1.$$

Pero, para el otro factor de (68.10) tenemos

$$\lim_{|z_k| \to \infty} |a_n||z_k|^n = +\infty.$$

Por consiguiente, la correlación (68.9) es válida.

TEOREMA 68.1 (*teorema fundamental del álgebra*). *Todo polinomio $f(z)$ de grado $n \geq 1$ con coeficientes complejos tiene por lo menos una raíz, que es, en el caso general, compleja.*

DEMOSTRACIÓN. Consideraremos un conjunto de toda clase de valores del módulo del polinomio $f(z)$. Puesto que $|f(z)| \geq 0$, este conjunto está acotado inferiormente. Del curso del análisis matemático sabemos que cualquier conjunto no vacío de números reales, acotado inferiormente, tiene la cota inferior exacta. Supongamos que para el conjunto de valores $|f(z)|$ esta cota es l. Esto significa que para cualquier número natural k se puede hallar tal número complejo z_k que

$$0 \leq |f(z_k)| - l \leq 2^{-k}.$$

De aquí se deduce que

$$\lim_{k \to \infty} |f(z_k)| = l. \qquad (68.11)$$

Si suponemos que la sucesión $\{z_i\}$ no está acotada, se podría elegir de ella una subsucesión infinitamente grande y, de acuerdo con el lema (68.3), la correlación (68.11) no podría verificarse. Por esta razón la sucesión $\{z_k\}$ es acotada. Elijamos de ésta una subsucesión convergente $\{z_{k_\nu}\}$ y supongamos que

$$\lim_{k_\nu \to \infty} z_{k_\nu} = z_0.$$

Conforme al corolario del lema 68.1, el módulo de un polinomio es una función continua. Por consiguiente,

$$|f(z_0)| = \lim_{k_\nu \to \infty} |f(z_{k_\nu})| = l.$$

Si $l \neq 0$, del lema 68.2 se desprende que existe tal número z_0', para el cual $|f(z_0')| < l$. Esto contradice a que l es la cota inferior exacta de los valores del módulo del polinomio, razón por la cual $l = 0$.

Así pues, hemos probado la existencia de un número complejo z_0 tal, que $|f(z_0)| = 0$, ó, lo que es igual,

$$f(z_0) = 0.$$

Esto significa que z_0 es la raíz del polinomio $f(z)$.

Ejercicios.

1. Demuéstrese que el conjunto de todas las raíces de n-ésimo grado del número complejo 1 forma un grupo conmutativo de multiplicación.
2. Demuéstrese que para que una sucesión de números complejos $\{z_k\}$ sea acotada, es necesario y suficiente que la sucesión $\{f(z_k)\}$ sea acotada por lo menos para un polinomio $f(z)$ de grado $n \geq 1$.
3. Demuéstrese que para cualquier polinomio $f(z)$ de grado $n \geq 1$ y todo número complejo z_0 existe un número complejo h tal que $|f(z_0 + h)| > |f(z_0)|$.
4. Demuéstrese que todas las raíces del polinomio (65.1) se hallan dentro del anillo

$$\left(1 + \max_{k>0}\left|\frac{a_k}{a_0}\right|\right)^{-1} \leq |z| \leq \left(1 + \max_{0 \leq k < n}\left|\frac{a_k}{a_n}\right|\right).$$

5. Rigiéndose por el mismo esquema que se ha usado para los números complejos, trátese de demostrar que el campo de números reales es algebraicamente cerrado. ¿En qué lugar la demostración no tiene analogía?

§ 69. Corolarios del teorema fundamental

Del teorema fundamental se deduce toda una serie de corolarios. Consideraremos los más importantes de ellos.

El polinomio $f(z)$ de grado $n \geq 1$ con coeficientes complejos tiene por lo menos una raíz z_1. Por esto, $f(z)$ puede ser descompuesto así:

$$f(z) = (z - z_1)\,\varphi(z),$$

donde $\varphi(z)$ es un polinomio de grado $n - 1$. Los coeficientes del polinomio $\varphi(z)$ son, como antes, números complejos. Por consiguiente, $\varphi(z)$ tiene la raíz z_2 (siempre que $n \geq 2$) y

$$\varphi(z) = (z - z_2)\,\psi(z),$$

de donde se deduce que

$$f(z) = (z - z_1)(z - z_2)\,\psi(z).$$

Continuando este proceso, obtendremos la descomposición del polinomio en un producto de factores lineales:

$$f(z) = b\,(z - z_1)(z - z_2)\dots(z - z_n),$$

donde b es un cierto número. Abriendo los paréntesis en el segundo miembro de la descomposición obtenida y comparando los coeficientes de las potencias con los coeficientes a_i del polinomio $f(z)$, llegamos a la conclusión de que $b = a_n$.

Entre los números $z_1, z_2, \dots, z_n$ pueden haber números iguales. Supongamos, para simplificar, que $z_1, \dots, z_r$ son distintos dos a dos, mientras que cada uno de los números $z_{r+1}, \dots, z_n$ es igual a uno de los primeros. En este caso el polinomio $f(z)$ puede escribirse en la forma:

$$f(z) = a_n (z - z_1)^{k_1} (z - z_2)^{k_2} \dots (z - z_r)^{k_r}, \qquad (69.1)$$

donde $z_i \neq z_j$ cuando $i \neq j$, y, además,

$$k_1 + k_2 + \dots + k_r = n.$$

La descomposición (69.1) se denomina *descomposición canónica del polinomio $f(z)$ en factores.*

Para el polinomio $f(z)$ la descomposición canónica es única, salvo el orden en que se disponen los factores. En efecto, supongamos que a la par con la descomposición (69.1) existe otra descomposición canónica, por ejemplo

$$f(z) = a_n (z - v_1)^{l_1} (z - v_2)^{l_2} \dots (z - v_m)^{l_m}.$$

En este caso será verídica la igualdad

$$(z - z_1)^{k_1} (z - z_2)^{k_2} \dots (z - z_r)^{k_r} = (z - v_1)^{l_1} (z - v_2)^{l_2} \dots (z - v_m)^{l_m}. \qquad (69.2)$$

Observemos que la totalidad de los números $z_1, \dots, z_r$ debe coincidir con la de los números $v_1, \dots, v_m$. Si, por ejemplo, z_1 no es igual a ninguno de los números $v_1, \dots, v_m$, entonces, al sustituir

$x = \alpha_1$ en (69.2), obtendremos cero en el primer miembro de la igualdad, mientras que en el segundo miembro, un número diferente de cero. Así pues, si existen dos descomposiciones canónicas del polinomio $f(x)$, la igualdad (69.2) sólo puede tener esta forma:

$$(x-\alpha_1)^{k_1}(x-\alpha_2)^{k_2}\dots(x-\alpha_r)^{k_r} = (x-\alpha_1)^{l_1}(x-\alpha_2)^{l_2}\dots(x-\alpha_r)^{l_r}.$$

Supongamos, por ejemplo, que $k_1 \neq l_1$ y sea, para concretar $k_1 > l_1$. Al dividir ambos miembros de la última igualdad por el mismo divisor $(x-\alpha_1)^{l_1}$, llegamos a que

$$(x-\alpha_1)^{k_1-l_1}(x-\alpha_2)^{k_2}\dots(x-\alpha_r)^{k_r} = (x-\alpha_2)^{l_2}\dots(x-\alpha_r)^{l_r}.$$

Otra vez, al sustituir aquí $x = \alpha_1$, vemos que en el primer miembro de la igualdad figura cero y en el segundo miembro, un número diferente de cero. De este modo, la unicidad de la descomposición canónica queda demostrada.

Si en la descomposición canónica (69.1) $k_1 = 1$, entonces la raíz α_1 se llama *simple*; en cambio si $k_1 > 1$, entonces la raíz α_1 se denomina *múltiple*. El número k_1 lleva el nombre de *multiplicidad* de la raíz α_1. Ahora podemos enunciar una deducción muy importante:

Todo polinomio de grado $n \geqslant 1$ con coeficientes complejos tiene n raíces, si cada una de las raíces se cuenta tantas veces cual es su multiplicidad.

Un polinomio de grado nulo no tiene raíces. El único polinomio que tiene un número tan grande como se quiera de raíces distintas dos a dos es el polinomio nulo. Haciendo uso de lo citado, podemos enunciar la deducción siguiente:

Si dos polinomios $f(x)$ y $g(x)$ de grado no superior a n tienen valores iguales para más de n diferentes valores del argumento, todos los coeficientes correspondientes de dichos polinomios serán iguales entre sí.

En efecto, el polinomio $f(x) - g(x)$ tiene, por hipótesis, más de n raíces. Pero su grado no es superior a n, por lo cual $f(x) - g(x) = 0$.

Así pues, un polinomio $f(x)$, cuyo grado no es superior a n, se determina completamente mediante sus valores para cualesquiera $n+1$ diferentes valores del argumento. Esto permite restablecer el polinomio por medio de sus valores. No es difícil señalar la forma explícita de este polinomio "restaurador". Si el polinomio $f(x)$ toma, para los valores del argumento $\alpha_1, \dots, \alpha_{n+1}$, los valores $f(\alpha_1), \dots, f(\alpha_{n+1})$, entonces

$$f(x) = \sum_{i=1}^{n+1} f(\alpha_i)\frac{(x-\alpha_1)\dots(x-\alpha_{i-1})(x-\alpha_{i+1})\dots(x-\alpha_{n+1})}{(\alpha_i-\alpha_1)\dots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\dots(\alpha_i-\alpha_{n+1})}.$$

Está claro que el grado del polinomio en el segundo miembro no es superior a n, y en los puntos $x = \alpha_i$ el polinomio toma los valores $f(\alpha_i)$. Un polinomio construido de esta manera se denomina *polinomio de interpolación de Lagrange*.

Examinemos ahora un polinomio $f(x)$ de grado n y supongamos que $\alpha_1, \alpha_2, \ldots, \alpha_n$ son sus raíces escritas tantas veces cual es la multiplicidad de cada una de ellas. En este caso

$$f(x) = a_n (x - \alpha_1)(x - \alpha_2) \ldots (x - \alpha_n).$$

Multiplicando las expresiones encerradas entre paréntesis que figuran en el segundo miembro, reduciendo los términos semejantes y comparando los coeficientes obtenidos con los de la expresión (68.3), podemos deducir las siguientes igualdades:

$$\begin{aligned}
a_{n-1}/a_n &= -(\alpha_1 + \alpha_2 + \ldots + \alpha_n),\\
a_{n-2}/a_n &= +(\alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_1\alpha_n + \ldots + \alpha_{n-1}\alpha_n),\\
a_{n-3}/a_n &= -(\alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_2\alpha_4 + \ldots + \alpha_{n-2}\alpha_{n-1}\alpha_n),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
a_1/a_n &= (-1)^{n-1}(\alpha_1\alpha_2 \ldots \alpha_{n-1} + \ldots + \alpha_2\alpha_3 \ldots \alpha_n),\\
a_0/a_n &= (-1)^n \alpha_1\alpha_2 \ldots \alpha_n.
\end{aligned}$$

Estas igualdades se llaman *fórmulas de Viète* y expresan los coeficientes de un polinomio en términos de sus raíces.

En el segundo miembro de la k-ésima igualdad figura la suma de toda clase de productos de k raíces tomada con el signo más o el signo menos, en dependencia de si es par o impar el número k.

En lo sucesivo nos serán útiles algunos corolarios del teorema fundamental del álgebra referentes a los polinomios con coeficientes reales. Supongamos que el polinomio

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n$$

con coeficientes reales tiene una raíz compleja v (pero no real), es decir,

$$a_0 + a_1 v + \ldots + a_n v^n = 0.$$

La última igualdad no se violará, si todos los números en ella se sustituyen por los números complejos conjugados. No obstante, los coeficientes $a_0, \ldots, a_n$ y el número 0, siendo números reales, quedarán en esta sustitución intactos. Por ello

$$a_0 + a_1\overline{v} + \ldots + a_n\overline{v}^n = 0,$$

es decir, $f(\overline{v}) = 0$.

De este modo, si un número complejo (pero no real) v es una raíz del polinomio $f(x)$ con coeficientes reales, entonces el número complejo conjugado $\overline{v}$ también será raíz de $f(x)$.

De aquí se deduce que el polinomio $f(x)$ se dividirá por un trinomio de segundo grado

$$\varphi(x) = (x - v)(x - \overline{v}) = x^2 - (v + \overline{v})x + v\overline{v},$$

cuyos coeficientes son reales. Haciendo uso de este hecho, demostremos que las raíces v y $\bar{v}$ son de la misma multiplicidad.

Supongamos que la multiplicidad de ellas es k y l, respectivamente, y, por ejemplo, $k > l$. En este caso $f(z)$ se divide por el l-ésimo grado del polinomio $\varphi(z)$, es decir,

$$f(z) = \varphi_0^l(z) \cdot q(z).$$

El polinomio $q(z)$, representando en sí un cociente de dos polinomios con coeficientes reales, tiene también coeficientes reales. Por hipótesis, el número v debe ser la raíz de multiplicidad $(k - l)$ de este polinomio que no tiene raíz igual a $\bar{v}$. De acuerdo con lo demostrado anteriormente, esto no es posible, razón por la cual $k = l$. De esta modo, todas las raíces complejas de cualquier polinomio de coeficientes reales son complejas conjugadas dos a dos. De la unicidad de la descomposición canónica proviene la siguiente deducción.

Todo polinomio con coeficientes reales puede ser representado, salvo el orden en que se disponen los factores, de un modo único en forma del producto de un coeficiente mayor y los polinomios con coeficientes reales. Los últimos polinomios tienen sus coeficientes mayores equivalentes a uno y son lineales, si corresponden a las raíces reales, y cuadráticos, cuando corresponden a un par de raíces complejas conjugadas.

Por fin, viene la deducción más importante, en aras de la cual, en esencia, se demostraba el teorema fundamental del álgebra. Supongamos que el operador lineal A actúa en un espacio complejo. Los valores propios de este operador, y sólo ellos, son las raíces del polinomio característico. Conforme al teorema fundamental, el operador A tiene por lo menos un valor propio λ. Por consiguiente, *Todo operador lineal que actúa en un espacio complejo lineal tiene por lo menos un vector propio.*

Ha de ser observado que si el operador A actúa en un espacio real y racional dicha deducción *ya no es justa.*

En relación con los valores propios se empleará la misma terminología que se ha aplicado respecto a las raíces del polinomio. En particular, un valor propio se llamará simple, si es una raíz simple del polinomio característico y se denominará múltiple, en el caso contrario. Se llamará multiplicidad del valor propio λ la multiplicidad de λ, al intervenir éste en calidad de raíz del polinomio característico.

Ejercicios.

1. Demuéstrese que si el número complejo $a \neq 0$, para todo número natural n sólo existen n números complejos distintos, cuya n-ésima potencia es igual a a.

2. ¿De que modo están ligadas entre sí las raíces de los polinomios $f(z)$ y $f(z - a)$, donde a es un número complejo?

3. Supongamos que el polinomio $f(z)$ con coeficientes reales, de grado no superior a n tome los valores iguales para $n+1$ distintos valores del argumento. Demuéstrese que $f(z)$ es un polinomio de grado nulo.

4. Demuéstrese que cualquier polinomio de grado impar, con coeficientes reales, tiene por lo menos una raíz real.

5. Demuéstrese que el polinomio $f(z)$ tiene por lo menos una raíz en cada uno de los dos dominios

$$|z| \leqslant \sqrt[n]{\left|\frac{a_0}{a_n}\right|}, \qquad |z| \geqslant \sqrt[n]{\left|\frac{a_0}{a_n}\right|}.$$

6. Demuéstrese que el operador A es de estructura simple, cuando, y sólo cuando, a todo valor propio le corresponden tantos vectores propios linealmente independientes cual es la multiplicidad de λ.

CAPITULO 9 ESTRUCTURA DEL OPERADOR LINEAL

§ 70. Subespacios invariantes

Las investigaciones que siguen se realizarán bajo el supuesto de que el operador lineal viene dado en un espacio complejo X. Como ya se ha indicado antes, dicha suposición asegura que para cualquier operador lineal existe al menos un vector propio.

El subespacio L de un espacio lineal X se denomina *invariante* respecto del operador A, si para todo vector x de L su imagen Ax pertenece también a L.

Todo operador lineal tiene al menos dos subespacios invariantes *triviales*: el subespacio nulo y todo el espacio X. Significación esencial sólo tienen los subespacios invariantes no triviales. A los subespacios semejantes pertenecen, por ejemplo, los subespacios propios. Como en un espacio lineal complejo cualquier operador tiene, a ciencia cierta, por lo menos un vector propio, cualquier operador en tal espacio tiene obligatoriamente al menos un subespacio invariante no trivial.

Es fácil comprobar que para todo operador A el campo de valores T_A y el núcleo N_A serán subespacios invariantes. Estos subespacios son triviales cuando, y sólo cuando, el operador A sea regular o nulo.

Si L es un subespacio invariante, existen varios métodos por medio de los cuales se puede construir un subespacio complementario M tal que $X = L \dotplus M$. Sin embargo, entre estos subespacios complementarios *puede no haber ninguno* que sea invariante. En cambio, si existe aunque sea un solo subespacio complementario invariante, podemos hablar sobre la descomposición del espacio en una suma directa de subespacios invariantes.

El conocimiento de algún subespacio invariante y más aún, de la descomposición del espacio en una suma directa de subespacios invariantes permite construir una base en la cual la matriz del operador tenga la forma más simple. Supongamos que el operador A tiene en el espacio m-dimensional X un subespacio invariante L de dimensión n. Elijamos en X una base $e_1, e_2, \ldots, e_m$ de tal modo

que sus primeros n vectores pertenezcan a L. En este caso las imágenes $Ae_1, \ldots, Ae_n$ de los vectores $e_1, \ldots, e_n$ pertenecerán a L y se podrá descomponerlas según los vectores $e_1, \ldots, e_n$, considerando los últimos como vectores de la base L. Por consiguiente,

$$Ae_1 = a_{11}e_1 + a_{21}e_2 + \ldots + a_{n1}e_n,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Ae_n = a_{1n}e_1 + a_{2n}e_2 + \ldots + a_{nn}e_n.$$

Recordemos que los elementos de las columnas de la matriz de un operador coinciden con las coordenadas de las imágenes de los vectores de la base. Por ello la matriz A_e del operador A en la base $e_1, e_2, \ldots, e_m$ tendrá por expresión:

$$A_e = \left(\begin{array}{ccc|ccc} a_{11} & \ldots & a_{1n} & a_{1,n+1} & \ldots & a_{1m} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} & a_{n,n+1} & \ldots & a_{nm} \\ \hline 0 & \ldots & 0 & a_{n+1,n+1} & \ldots & a_{n+1,m} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & a_{m,n+1} & \ldots & a_{mm} \end{array}\right).$$

Por regla general, las matrices de este tipo se escriben en la así llamada forma *celular*. A saber,

$$A_e = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}. \qquad (70.1)$$

Aquí, A_{11} es una matriz cuadrática de orden n, A_{22} es una matriz cuadrática de orden $m - n$, 0 es una matriz nula de dimensiones $(m - n) \times n$, y A_{12}, la matriz de dimensiones $n \times (m - n)$.

Supongamos ahora que el espacio X está descompuesto en suma directa de los subespacios invariantes L y M. Elijamos en X una base $e_1, e_2, \ldots, e_m$ de modo tal que los primeros n vectores de la base pertenezcan a L y los demás $m - n$ vectores pertenezcan a M. En este caso las imágenes $Ae_1, \ldots, Ae_n$ pueden descomponerse sólo según los vectores $e_1, \ldots, e_n$, mientras que las imágenes $Ae_{n+1}, \ldots, Ae_m$, sólo según los vectores $e_{n+1}, \ldots, e_m$. La matriz A_{12} en (70.1) será, evidentemente, nula. Por esta razón la matriz A_e del operador A en la base considerada tendrá una forma aún más simple. A saber,

$$A_e = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}.$$

Supongamos que la operación del operador A se investiga sólo en los vectores del subespacio invariante L. Si $x \in L$, se tiene $Ax \in L$. Por consiguiente, podemos considerar que el operador A genera en L

otro operador $A \mid L$ definido por la igualdad

$$(A \mid L)\, x = Ax$$

para todo $x \in L$. El operador $A \mid L$ se denomina *operador inducido, generado por el operador A*. Con relación al operador $A \mid L$, el operador A se llama *generador*. Por ser el operador A lineal, el operador inducido será también lineal. El operador inducido $A \mid L$ coincide con el operador generador A en el subespacio invariante L y no está definido fuera de L. De este modo, dichos operadores se diferencian, principalmente, por el campo de definición.

A pesar de que la introducción del operador inducido parece ser algo artificial, éste representa en sí un aparato auxiliar muy cómodo en la ejecución de las más diversas investigaciones. Por ejemplo, un operador inducido, al igual que cualquier otro operador lineal, tiene por lo menos un vector propio. Pero, como dicho operador coincide con el operador generador en su campo de definición, esto significa que:

Todo operador lineal tiene en cualquier subespacio invariante por lo menos un vector propio.

Si un espacio está descompuesto en una suma directa de r subespacios invariantes, el operador lineal tiene por lo menos r vectores propios linealmente independientes.

Está claro que todo valor propio y todo vector propio de un operador inducido son, respectivamente, el valor propio y el vector propio del operador generador. Resulta menos evidente el

TEOREMA 70.1 *El polinomio característico de un operador inducido engendrado en un subespacio no trivial es el divisor del polinomio característico del operador generador.*

DEMOSTRACIÓN. Supongamos que el operador inducido $A \mid L$ está definido en el subespacio invariante L. Elijamos nuevamente en el espacio X una base $e_1, \ldots, e_m$ de tal modo que los vectores $e_1, \ldots, e_n$ compongan la base en L. Si la matriz del operador generador es A_e de (70.1), la matriz del operador inducido $A \mid L$ es A_{11} de (70.1). El polinomio característico para el operador A es igual a $\det(\lambda E - A_e)$; para el operador $A \mid L$ el mismo es igual a $\det(\lambda E - A_{11})$. Aplicando el teorema de Laplace para el desarrollo del determinante $\det(\lambda E - A_e)$ por las primeras n columnas, encontramos

$$\det(\lambda E - A_e) = \det\begin{pmatrix} \lambda E - A_{11} & -A_{12} \\ 0 & \lambda E - A_{22} \end{pmatrix} = \det(\lambda E - A_{11}) \times$$
$$\times \det(\lambda E - A_{22}).$$

La igualdad obtenida significa precisamente la validez de la afirmación del teorema.

La determinación de todos los valores propios del operador A se reduce a la búsqueda de todas las raíces del polinomio caracte-

rístico. Si el operador A tiene un subespacio invariante no trivial, entonces, de acuerdo con el teorema 70.1, la tarea puede ser reducida a la búsqueda de todas las raíces de dos polinomios de grados inferiores. Si el operador inducido tiene por sí mismo un subespacio invariante no trivial, el proceso de descomposición del polinomio característico en factores puede ser continuado.

Ejercicios.

1. Demuéstrese que la suma y la intersección de los subespacios invariantes son unos subespacios invariantes.
2. Demuéstrese que si el operador A es regular, entonces cualquier operador inducido es también regular.
3. Demuéstrese que si el operador A es de estructura simple, cualquier operador inducido es también de estructura simple.
4. ¿En qué caso el subespacio invariante de un operador de estructura simple será la suma directa de los subespacios propios?
5. Demuéstrese que si aunque sea uno de los subespacios invariantes del operador A está privado de un subespacio invariante complementario, A no puede ser operador de estructura simple.
6. Demuéstrese que si el operador A es de estructura simple, el campo de valores y el núcleo están privados de vectores no nulos comunes.
7. Demuéstrese que si un subespacio es invariante respecto del operador A, será invariante también respecto del operador $\alpha_0 E + \alpha_1 A + \ldots + \alpha_p A^p$.

§ 71. Polinomio operacional

Uno de los métodos más importantes para construir subespacios invariantes de un operador lineal consiste en el empleo de los polinomios con coeficientes complejos.

Supongamos que el operador lineal A actúa en el espacio complejo X. Elijamos un polinomio arbitrario

$$\varphi(z) = \alpha_0 + \alpha_1 z + \ldots + \alpha_p z^p$$

con coeficientes complejos y consideremos un operador lineal

$$\varphi(A) = \alpha_0 E + \alpha_1 A + \ldots + \alpha_p A^p.$$

Este operador actúa en el espacio X y se denomina *polinomio operacional* o *polinomio del operador A*.

Fijemos el operador A y construyamos el conjunto de todos los polinomios operacionales que dependen de A. Puesto que el conjunto de todos los polinomios es un anillo conmutativo, lo será también el conjunto de todos los polinomios operacionales. En particular, de aquí se deduce que

$$\varphi(A)\, A = A \varphi(A)$$

para cualquier polinomio $\varphi(z)$. *La conmutatividad de un anillo de polinomios operacionales desempeña un papel exclusivamente importante en todas las investigaciones a seguir.*

Es fácil mostrar que el campo de valores T_φ de todo polinomio operacional $\varphi(A)$ es un subespacio invariante para el operador A. En efecto, sea $x \in T_\varphi$. Esto significa que $x = \varphi(A)y$ para cierto $y \in X$. Por ser A y $\varphi(A)$ conmutables, tenemos

$$Ax = A\varphi(A)y = \varphi(A)(Ay).$$

Por consiguiente, el vector Ax es el resultado de aplicar el operador $\varphi(A)$ al vector $Ay \in X$, es decir, $Ax \in T_\varphi$.

El núcleo N_φ del polinomio operacional $\varphi(A)$ es también un subespacio invariante para el operador A. Si $x \in N_\varphi$, entonces $\varphi(A)x = 0$, pero en este caso

$$\varphi(A)(Ax) = A(\varphi(A)x) = A(0) = 0.$$

Ya se ha observado anteriormente que cualquier subespacio invariante contiene por lo menos un vector propio del operador. Ahora podemos enunciar una afirmación más exacta. A saber,

Si el valor propio del operador A es (no es) una raíz del polinomio $\varphi(z)$, todos los vectores propios del operador A, correspondientes a dicho valor propio, pertenecen al núcleo (campo de valores) del operador $\varphi(A)$.

Efectivamente, sea x un vector propio del operador A correspondiente al valor propio λ. En los ejercicios del § 65 se subrayaba que el vector x es también propio para el operador $\varphi(A)$, pero corresponde al valor propio $\varphi(\lambda)$. Por consiguiente, $\varphi(A)x = \varphi(\lambda)x$. Si λ es una raíz del polinomio $\varphi(z)$, entonces $\varphi(\lambda) = 0$ y el vector x pertenece al núcleo del operador $\varphi(A)$. En cambio si $\varphi(\lambda) \neq 0$, el vector $\varphi(\lambda)x$ será no nulo y el vector x pertenecerá al campo de valores del operador $\varphi(A)$.

No podemos demostrar que cualquier subespacio invariante del operador A es o bien un campo de valores o bien un núcleo de cierto polinomio operacional. Esta afirmación, por lo general, no es cierta, lo cual se afirma en el ejemplo del operador idéntico. Cualquiera que sea el polinomio $\varphi(z)$, se tiene $\varphi(E) = \varphi(1)E$, por lo cual el operador $\varphi(E)$ es o bien nulo o bien regular. Por consiguiente, el campo de valores y el núcleo para $\varphi(E)$ representan siempre unos subespacios triviales. En cuanto al subespacio invariante del operador E, lo constituirá cualquier subespacio. Con todo esto, cada subespacio invariante del operador A está ligado de modo determinado a los polinomios operacionales de A. Es válido el

TEOREMA 71.1. *Sea L un subespacio invariante arbitrario del operador A. Si todos los valores propios del operador, inducido sobre L, son las raíces del polinomio $\varphi(z)$, entonces L está contenido en los núcleos de los operadores $\varphi^k(A)$ para todas las potencias k, enteras positivas y suficientemente grandes.*

DEMOSTRACIÓN. Designaremos mediante $T'_1, T'_2, \ldots$ los campos de valores de los operadores inducidos sobre L con ayuda de los polinomios operacionales $\varphi^k(A)$ para $k = 1, 2, \ldots$. El operador

$\varphi(A)$ es degenerado en L, puesto que en su núcleo están contenidos por lo menos todos los vectores propios del operador A pertenecientes a L. Por ello, $T'_1 \subset L$, $\dim T'_1 < \dim L$. El subespacio T'_1 es invariante respecto de A. Si T'_1 es no nulo, entonces, de acuerdo con el teorema (70.1), el polinomio característico del operador inducido sobre T'_1 con ayuda de A es el divisor del polinomio característico del operador inducido sobre L con ayuda de A. Por consiguiente, todos los valores propios del operador, inducido sobre T'_1, son también las raíces del polinomio $\varphi(z)$. Pero de aquí se desprende nuevamente que $T'_2 \subset T'_1$, $\dim T'_2 < \dim T'_1$, etc. Las dimensiones de los subespacios T'_1, T'_2, ... no pueden decrecer ilimitadamente. Por esta razón, a partir de cierto número k, dichos subespacios irán quedando nulos lo que es testimonio de la validez de la afirmación del teorema.

Las investigaciones efectuadas permiten establecer un hecho muy importante concerniente a la existencia de los subespacios invariantes no triviales.

TEOREMA 11.2 *Todo operador lineal A, que actúa en el espacio complejo m-dimensional X, tiene por lo menos un subespacio invariante de dimensión $m - 1$.*

DEMOSTRACIÓN. El operador A tiene por lo menos un vector propio x. Supongamos que dicho vector corresponde al valor propio λ. De acuerdo con lo demostrado, el campo de valores T_λ del operador $A - \lambda E$ representa un subespacio invariante del operador A. Pero, puesto que el operador $A - \lambda E$ es degenerado, la dimensión del subespacio T_λ no es superior a $m - 1$.

Consideraremos ahora cualquier subespacio L de dimensión $m - 1$ en el que está íntegramente contenido el subespacio T_λ. Todo vector del espacio X se transforma por el operador $A - \lambda E$ en cierto vector de T_λ. Por eso todo vector de L pasa a ser, de nuevo, un vector de L. De este modo, L es el subespacio invariante respecto de $A - \lambda E$, y, por supuesto, invariante respecto de A. El teorema queda demostrado.

Ejercicios

1. Sea A un operador de diferenciación que actúa en un espacio real de polinomios, dimensión finita. ¿Qué representa en sí el operador $\varphi(A)$ para el polinomio $\varphi(z)$ con coeficientes reales?

2. Sea $\varphi(z)$ un polinomio característico del operador inducido generado por el operador A en el subespacio invariante N. Demuéstrese que N pertenece al núcleo del operador $\varphi^k(A)$ para cierto k entero y positivo.

3. Demuéstrese que si todos los valores propios del operador A son las raíces del polinomio $\varphi(z)$, entonces $\varphi^k(A) = 0$ para cierto k entero y positivo.

4. Demuéstrese que un anillo de polinomios operacionales generados por un operador cualquiera tiene divisores de cero.

5. Demuéstrese que si el operador A es de estructura simple, el operador $\varphi(A)$ será también de estructura simple. ¿Será cierta la afirmación inversa?

§ 72. Forma triangular

Ahora podemos resolver el problema de reducción de la matriz de un operador a una de las formas más sencillas, esto es, a la así llamada *forma triangular*.

TEOREMA 72.1. *Para cualquier operador lineal A que actúa en un espacio m-dimensional X existen tales subespacios invariantes L_p de dimensión p, $p = 0, 1, \dots, m-1$, que*

$$L_0 \subset L_1 \subset \dots \subset L_{m-1} \subset L_m.$$

DEMOSTRACIÓN. La existencia de los subespacios L_0 y L_m es evidente. Conforme a lo demostrado más arriba, el operador A tiene subespacio invariante L_{m-1} de dimensión $m-1$.

Consideraremos en el subespacio L_{m-1} un operador inducido. Al igual que cualquier otro operador prefijado en L_{m-1}, el operador inducido tiene el subespacio invariante L_{m-2} de dimensión $m-2$. Pero un subespacio que es invariante para el operador inducido será también invariante para el operador generador A. De este modo, la existencia del subespacio L_{m-2} está demostrada. Al considerar un operador inducido en el subespacio L_{m-2}, veremos que de modo análogo se demuestra la existencia del subespacio L_{m-3}, etc.

Este teorema es interesante, en primer lugar, por su interpretación matricial. Construyamos la base $e_1, e_2, \dots, e_m$ del espacio X, haciendo uso de los subespacios invariantes L_p. A título de vector e_1 tomemos cualquier vector no nulo de L_1, a título de vector e_2 tomemos cualquier vector no nulo de L_2, no perteneciente a L_1, y, en general, a título de vector e_p tomemos cualquier vector no nulo de L_p, no perteneciente a L_{p-1}. Consideremos la matriz A_e del operador A en esta base. Puesto que e_j pertenece a L_j, mientras que L_j es invariante respecto de A, entonces el vector Ae_j debe representar una combinación lineal sólo de los vectores $e_1, e_2, \dots, e_j$. Quiere decir que en la descomposición

$$Ae_j = a_{1j}e_1 + a_{2j}e_2 + \dots + a_{mj}e_m$$

el coeficiente de e_i debe ser igual a cero para cualesquiera $i > j$. Por consiguiente, la matriz del operador A tiene la forma

$$A_e = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,m-1} & a_{1m} \\ 0 & a_{22} & \dots & a_{2,m-1} & a_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{m-1,m-1} & a_{m-1,m} \\ 0 & 0 & \dots & 0 & a_{mm} \end{pmatrix}$$

donde $a_{ij} = 0$ para $i > j$.

Una matriz, todos los elementos de la cual, dispuestos por debajo (por arriba) de la diagonal principal, son nulos, se denomina *matriz*

triangular derecha (izquierda). En el lenguaje de las matrices el resultado obtenido significa que toda matriz cuadrada es semejante a una matriz triangular derecha.

La forma triangular de una matriz es de amplio uso en la demostración de los más diversos hechos concernientes a los operadores lineales. Esto se debe, principalmente, a la siguiente propiedad de la misma:

Si el operador A tiene en cierta base la matriz triangular A_e, los elementos diagonales de la matriz A_e coinciden con los valores propios del operador A incluso cuando se toma en consideración la multiplicidad de éstos.

En efecto, al aplicar el teorema de Laplace, encontramos que el polinomio característico de la matriz A_e es igual a

$$\det(\lambda E - A_e) = \prod_{i=1}^{n} (\lambda - a_{ii}),$$

de donde proviene que la afirmación enunciada es cierta.

Una parte considerable de la ulterior teoría de los operadores lineales se dedica al perfeccionamiento del resultado recién obtenido acerca de la reducción de la matriz de un operador a la forma triangular. La forma más simple que puede tener la matriz de un operador es diagonal. Como se sabe, a esta forma pueden reducirse sólo las matrices de los operadores de estructura simple. No obstante, para los operadores de estructura no simple la forma triangular tampoco es la más sencilla.

Ejercicios

1. Demuéstrese que cualquier matriz cuadrada es semejante a una matriz triangular izquierda.
2. Demuéstrese que un conjunto de matrices triangulares de un mismo orden y de una misma denominación forma un anillo.
3. Demuéstrese que un conjunto de matrices triangulares regulares de un mismo orden y de la misma denominación forma un grupo.
4. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los valores propios del operador A escritos por orden tomando en consideración su multiplicidad. Demuéstrese que, tomando en consideración la multiplicidad, los números $\varphi(\lambda_1), \varphi(\lambda_2), \ldots, \varphi(\lambda_n)$ serán valores propios del operador $\varphi(A)$, cualquiera que sea el polinomio $\varphi(z)$.
5. Demuéstrese que si todos los elementos diagonales de una matriz triangular A de orden n son nulos, entonces $A^n = 0$.
6. Supongamos que una matriz triangular es semejante a la diagonal. Demuéstrese que la matriz de la transformación de semejanza puede ser elegida triangular de la misma denominación.

§ 73. Suma directa de los operadores

Un operador lineal cuyos valores propios son todos iguales es, en cierto sentido, una exclusión. No obstante, mostraremos que todo operador lineal puede ser compuesto precisamente de operadores de este tipo.

Supongamos que el espacio X está representado en forma de la suma directa de los subespacios L y M. Fijemos en el subespacio L cierto operador B y en el subespacio M, el operador C. Para todo vector $x \in X$ tiene lugar la descomposición única

$$x = x_L + x_M, \tag{73.1}$$

donde $x_L \in L$, $x_M \in M$.

Un operador A, definido mediante la igualdad

$$Ax = Bx_L + Cx_M,$$

se llama *suma directa* de los operadores B y C. Si uno de los subespacios L, M es trivial, la suma directa se denomina también trivial.

Es fácil comprobar que A es un operador lineal en X. Señalemos que dicho operador puede ser representado de un modo único en forma de la suma directa de los operadores definidos en los subespacios L y M. En efecto, para todo vector $x \in L$ se tiene $Ax = Bx$. Por analogía, $Ax = Cx$ para todo $x \in M$. Esto significa que el operador C coincide con el operador inducido $A \mid L$ y el operador C coincide con el $A \mid M$.

Consideremos ahora un operador arbitrario A en el espacio X. Si X queda descompuesto de tal o cual manera en la suma directa de los subespacios L y M, invariantes respecto del operador A, entonces el propio *operador A puede ser descompuesto en suma directa*. Efectivamente, construyamos los operadores inducidos $A \mid L$ y $A \mid M$. Al descomponer una vez más el vector arbitrario $x \in X$ en forma de la suma (73.1), obtendremos

$$Ax = (A \mid L)\, x_L + (A \mid M)\, x_M.$$

En este caso, en virtud del teorema 70.1, el polinomio característico del operador A es igual al producto de los polinomios característicos de los operadores inducidos $A \mid L$ y $A \mid M$.

El operador A puede descomponerse en una suma directa con ayuda de cualquier polinomio operacional $\varphi(A)$. Designemos con N_k el núcleo del operador $\varphi^k(A)$. Esto es un subespacio invariante respecto de A y, evidentemente, $N_1 \subset N_2 \subset \ldots$ Demostremos primeramente que si $N_k = N_{k+1}$ para cierto k, entonces $N_k = N_p$ para todo $p > k$. En efecto, elijamos cualquier vector $x \in N_p$, entonces $\varphi^p(A)\, x = 0$. Al escribir esta igualdad en la forma $\varphi^{k+1}(A) \times \times (\varphi^{p-k-1}(A)\, x) = 0$, concluimos que el vector $\varphi^{p-k-1}(A)\, x \in N_{k+1}$. En virtud de la igualdad $N_k = N_{k+1}$, este mismo vector pertenece a N_k. Por consiguiente,

$$\varphi^k(A)\,(\varphi^{p-k-1}(A)\, x) = \varphi^{p-1}(A)\, x = 0,$$

es decir, el vector $x \in N_{p-1}$. La validez de la afirmación enunciada se establece ahora por inducción según p.

El espacio X, donde actúa el operador A, es de dimensión finita, por lo cual las dimensiones de los subespacios N_k no pueden crecer

ilimitadamente. Sea q un número positivo entero mínimo, para el cual $N_q = N_{q+1}$. Designemos mediante T_k el campo de valores del operador $\varphi^k(A)$ y examinemos cualquier vector común x de los subespacios T_q y N_q. Tenemos $\varphi^q(A)x = 0$ y $x = \varphi^q(A)y$ para cierto vector $y \in X$. De aquí proviene que $\varphi^{2q}(A)y = 0$, es decir, $y \in N_{2q}$. Pero, de acuerdo con lo demostrado, se verifica la igualdad $N_q = N_{2q}$. Por esta razón $y \in N_q$, es decir, $x = \varphi^q(A)y = 0$.

Así pues, para los subespacios T_q y N_q sólo el vector nulo resulta ser común. En virtud de la fórmula (56.3), esto significa que $X = T_q \dotplus N_q$. Puesto que los subespacios T_q y N_q son invariantes, la posibilidad de descomponer al operador queda establecida.

Según se ha observado anteriormente, todos los vectores propios del operador A deben encontrarse en los subespacios T_q y N_q. Además, en N_q se encuentran aquellos de los vectores que corresponden a los valores propios, coincidentes con las raíces cualesquiera del polinomio $\varphi(z)$, mientras que en T_q se hallan aquellos vectores propios, para los cuales los valores propios correspondientes no coinciden con ninguna de las raíces de $\varphi(z)$. Puesto que a todo valor propio le corresponde aunque sea un solo vector propio, de estos razonamientos se deduce que:

Cada una de las raíces (ninguna de las raíces) del polinomio característico de un operador, inducido sobre N_q (T_q), es (no es) la raíz del polinomio $\varphi(z)$.

La característica definitiva de las descomposiciones de un operador en una suma directa, obtenida con ayuda de los polinomios operacionales, la proporciona el

TEOREMA 73.1. *Supongamos que el polinomio característico $f(z)$ del operador A está descompuesto en el producto de los polinomios $\varphi(z)$ y $\psi(z)$ que no tienen raíces comunes. En este caso el operador A puede descomponerse del modo único en la suma directa de los operadores B y C con polinomios característicos $\varphi(z)$ y $\psi(z)$.*

DEMOSTRACIÓN. Consideraremos la descomposición del operador A en una suma directa, obtenida con ayuda del polinomio $\varphi(z)$. Como el producto de los polinomios característicos de los operadores, que definen la suma directa, coincide con el polinomio característico $f(z)$, entonces la existencia de al menos una descomposición se deduce de las investigaciones realizadas más arriba.

Supongamos ahora que el espacio X está descompuesto de uno u otro modo en la suma directa de los espacios invariantes N y T. En este caso el operador inducido tiene en N un polinomio característico $\varphi(z)$ y el operador en T, el polinomio $\psi(z)$. Según el teorema 71.1, $N \subset N_k$ para cualesquiera k suficientemente grandes, por lo cual $N \subset N_q$. El operador $\varphi(A)$ es regular en T y, por consiguiente, el conjunto de imágenes de los vectores de T respecto de $\varphi(A)$ coincide con T. Pero esto precisamente significa que $T \subset T_k$

para todo h. Los subespacios N, T, como también N_q, T_q, forman en suma directa el espacio X. Por ello, las inclusiones $N \subset N_q$, $T \subset T_q$ pueden tener lugar sólo en el caso en que $N = N_q$, $T = T_q$. El teorema está demostrado.

Sea un operador A que actúa en el espacio m-dimensional X. Representemos el polinomio característico $f(z)$ del operador A en forma de la descomposición canónica

$$f(z) = (z - \lambda_1)^{k_1} (z - \lambda_2)^{k_2} \dots (z - \lambda_r)^{k_r}, \qquad (73.2)$$

donde $\lambda_1, \lambda_2, \dots, \lambda_r$ son unos valores propios distintos dos a dos y $k_1 + k_2 + \dots + k_r = m$. Examinemos los polinomios

$$(z - \lambda_1)^{k_1}, \ (z - \lambda_2)^{k_2}, \ \dots, \ (z - \lambda_r)^{k_r}.$$

Estos son los divisores del polinomio característico $f(z)$ y ningún par de ellos tiene raíces comunes. De acuerdo con el teorema 73.1, existen los subespacios invariantes $R_1, R_2, \dots, R_r$ tales que

$$X = R_1 + R_2 + \dots + R_r.$$

En estas circunstancias la dimensión del subespacio R_i es igual a k_i y el operador inducido en R_i tiene el polinomio característico $(z - \lambda_i)^{k_i}$.

El subespacio R_i se llama subespacio *radical* del operador A, correspondiente al valor propio λ_i. Los vectores de un subespacio radical se denominan *radicales*. De lo dicho se deduce que todo operador puede ser descompuesto en la suma directa de los operadores inducidos en los subespacios radicales.

El subespacio radical R_i coincide con el núcleo del operador $((A - \lambda_i E)^{k_i})^q$ para cierto q entero y positivo. Probemos que en el caso dado siempre se puede poner $q = 1$. Examinemos los operadores $(A - \lambda_i E)^p$ para $p = 1, 2, \dots$ Sea p_i un número mínimo para el cual el núcleo del operador $(A - \lambda_i E)^{p_i}$ coincide con el del operador $(A - \lambda_i E)^{p_i + 1}$. En este caso el subespacio radical R_i coincidirá con el núcleo del operador $(A - \lambda_i E)^{p_i}$. Como las dimensiones de los núcleos de los operadores $(A - \lambda_i E)^p$ para $p = 1, 2, \dots$ crecen de modo monótono y la dimensión del subespacio R_i es igual a k_i, resulta que $p_i \leqslant k_i$.

De este modo, R_i, correspondiente al valor propio λ_i de multiplicidad k_i, coincide, a ciencia cierta, con el núcleo del operador $(A - \lambda_i E)^{k_i}$.

TEOREMA 73.2. *(teorema de Cayley—Hamilton). Si $f(z)$ es un polinomio característico del operador A, entonces $f(A)$ es un operador nulo.*

DEMOSTRACIÓN. Representemos el polinomio característico en forma de la descomposición canónica (73.2). Puesto que el polinomio operacional $f(A)$ contiene el factor $(A - \lambda_i E)^{k_i}$ y cualesquiera

polinomios de un mismo operador son permutables, $f(A)x_i = 0$ para todo vector x_i de R_i. Elijamos ahora un vector arbitrario x y representémoslo en forma de la descomposición $x = x_1 + x_2 + \ldots + x_r$, donde $x_i \in R_i$. Ahora está claro que $f(A)x = 0$, es decir, $f(A)$ es un operador nulo.

Esta vez también representa un interés considerable la interpretación matricial de los resultados obtenidos. Formemos la base de un espacio como reunión sucesiva de cualesquiera bases de los subespacios radicales $R_1, R_2, \ldots, R_r$. Los subespacios radicales son invariantes y la suma directa de ellos coincide con X. Por esta razón, la matriz A_e del operador A tendrá en la base dada la así llamada forma *casi diagonal*

$$\begin{pmatrix} A_{11} & & & 0 \\ & A_{22} & & \\ & & \ddots & \\ 0 & & & A_{rr} \end{pmatrix}. \tag{73.3}$$

Toda matriz A_{ii} tiene el orden k_i y representa en sí una matriz del operador inducido sobre el subespacio R_i.

Ejercicios.

1. ¿Se podrá descomponer en una suma directa no trivial el operador de diferenciación dado en un espacio de polinomios de dimensión finita?

2. Demuéstrese que un sistema de vectores radicales, que corresponden dos a dos a los diferentes valores propios, es linealmente independiente.

3. Demuéstrese que si el operador A es regular, entonces $A^{-1} = \varphi(A)$ para cierto polinomio $\varphi(t)$.

4. Un operador A se denomina *nilpotente*, si $A^p = 0$ para cierto p entero y positivo. Demuéstrese que un operador es nilpotente cuando, y sólo cuando, todos sus valores propios son iguales a cero.

5. Sea $\varphi(t)$ un polinomio de grado inferior, para el cual $\varphi(A) = 0$. Demuéstrese que $\varphi(t)$ es un divisor para el polinomio característico del operador A.

§ 74. Forma de Jordan

La simplificación ulterior de la matriz de un operador en comparación con la forma casi diagonal (73.3) puede llevarse a cabo sólo a cuenta de la construcción especial de las bases de cada uno de los subespacios radicales. Por supuesto, las bases radicales se pueden escoger de un modo tal que cualquier matriz A_{ii}, en (73.3) sea triangular. No obstante, esta forma de la matriz de un operador tampoco es la más sencilla.

Volvamos al estudio más detallado de la estructura de subespacios radicales. Si es que $x \in R_i$, entonces $(A - \lambda_i E)^{k_i} x = 0$. Pero para cada vector concreto x es muy posible que se verifique la igualdad $(A - \lambda_i E)^m x = 0$ también con $m < k_i$. En particular, si x es

un vector propio correspondiente al valor propio múltiple λ_i, entonces $(A - \lambda_i E) x = 0$, aunque $k_i \geq 2$.

Se llama *altura* del vector radical x un número entero mínimo no negativo m, para el cual $(A - \lambda_i E)^m x = 0$.

Todos los vectores radicales correspondientes al valor propio λ_i tienen alturas no superiores a la multiplicidad de λ_i. Sin embargo, recordemos que en el caso general las alturas de los vectores radicales y las multiplicidades de los valores propios son dos conceptos diferentes. Así, por ejemplo, para un operador de estructura simple no existen, en general, vectores radicales cuya altura sea más que uno, independientemente de las multiplicidades de los valores propios.

Sea R_i un subespacio radical correspondiente al valor propio λ_i de multiplicidad k_i. Designemos mediante t la altura máxima de los vectores radicales de R_i. Está claro que $t \leq k_i$. Si el vector x es de altura k, el vector $(A - \lambda_i E) x$ tendrá la altura $k - 1$. Por esto, en el subespacio R_i hay vectores radicales de cualquier altura desde 0 hasta t.

Para todo $k \leq t$, designemos con H_k la totalidad de todos los vectores cuyas alturas no son superiores a k. Es fácil mostrar que H_k es un subespacio en R_i. Si $x, y \in H_k$, entonces $(A - \lambda_i E)^k x = (A - \lambda_i E)^k y = 0$. Pero en este caso con cualesquiera números α, β tenemos $(A - \lambda_i E)^k (\alpha x + \beta y) = 0$, es decir, $\alpha x + \beta y \in H_k$. Luego, es evidente que

$$0 = H_0 \subset H_1 \subset \ldots \subset H_{t-1} \subset H_t = R_i.$$

Las dimensiones de estos subespacios se designarán mediante m_k, $0 = m_0 < m_1 < \ldots < m_{t-1} < m_t = k_i$.

Sean $f_1, \ldots, f_{p_t}$ unos vectores arbitrarios linealmente independientes de H_t cuya cápsula lineal dé H_t en la suma directa con H_{t-1}. Está claro que éstos serán los vectores radicales de altura t, $p_t = m_t - m_{t-1}$ y no existe ninguna combinación lineal no nula de los vectores $f_1, \ldots, f_{p_t}$ que pertenezca a H_{t-1}. Consideraremos una totalidad de los vectores

$$\begin{array}{ccc} f_1, & \ldots, & f_{p_t} \\ (A-\lambda_i E) f_1, & \ldots, & (A-\lambda_i E) f_{p_t}, \\ (A-\lambda_i E)^2 f_1, & \ldots, & (A-\lambda_i E)^2 f_{p_t}, \\ \ldots & \ldots & \ldots \\ (A-\lambda_i E)^{t-1} f_1, & \ldots, & (A-\lambda_i E)^{t-1} f_{p_t}. \end{array} \qquad (74.1)$$

Probemos que estos vectores son linealmente independientes. En efecto, formemos su combinación lineal e igualemos ésta a cero. Aplicando a ambos miembros de la igualdad obtenida el operador $(A - \lambda_i E)^{t-1}$, llegamos a que la combinación lineal de los vectores $f_1, \ldots, f_{p_t}$ se transforma por el operador $(A - \lambda_i E)^{t-1}$ en un vector

nula, es decir, dicha combinación es un vector de H_{t-1}. Por consiguiente, los coeficientes de estos vectores deben ser nulos. Apliquemos ahora a la misma igualdad el operador $(A - \lambda_i E)^{t-2}$. De manera análoga nos convencemos de que los coeficientes de los vectores dispuestos en la segunda fila de (74.1) deben ser nulos, etc.

Hemos de notar que por ser elegidos los vectores $f_1, \dots, f_{p_t}$ no existe ninguna combinación lineal no nula de los vectores dispuestos en la i-ésima fila de (74.1) que pertenezca a H_{t-1}.

Completemos los vectores $(A - \lambda_i E) f_1, \dots, (A - \lambda_i E) f_{p_t}$ con tales vectores $f_{p_t+1}, \dots, f_{p_{t-1}}$ de H_{t-1} que toda esta totalidad sea linealmente independiente y que su cápsula lineal dé H_{t-1} en suma directa con H_{t-2}. Está claro que éstos serán vectores radicales de altura $t - 1$, $p_{t-1} = m_{t-1} - m_{t-2}$, y no existe ninguna combinación lineal no nula de los vectores dados que pertenezca a H_{t-2}. Construyamos nuevamente una totalidad de vectores

$$
\begin{array}{l}
f_{p_t+1}, \dots, f_{p_{t-1}}, \\
(A-\lambda_i E) f_{p_t+1}, \dots, (A-\lambda_i E) f_{p_{t-1}}, \\
\dots\dots\dots\dots\dots\dots \\
(A-\lambda_i E)^{t-2} f_{p_t+1}, \dots, (A-\lambda_i E)^{t-2} f_{p_{t-1}}.
\end{array}
\qquad (74.2)
$$

Con relación a la totalidad de los vectores $(A - \lambda_i E) f_1, \dots, (A - \lambda_i E) f_{p_t}, f_{p_t+1}, \dots, f_{p_{t-1}}$ pueden demostrarse las mismas afirmaciones que se han demostrado respecto de la totalidad de los vectores $f_1, \dots, f_{p_t}$ al sustituir, por supuesto, t por $t - 1$. Pasando del mismo modo a los subespacios $H_{t-2}, H_{t-3}, \dots, H_1$, obtendremos un sistema linealmente independiente de k_i vectores pertenecientes al subespacio radical R_i. Las tablas del tipo (74.1), (74.2) terminan con una tabla de una línea

$$
f_{p_2+1}, \dots, f_{p_1}. \qquad (74.3)
$$

Estos vectores pertenecen a H_1, es decir, son propios, $p_1 = m_1 - m_0$.

Dispondremos las tablas del tipo (74.1)—(74.3) sucesivamente de izquierda a derecha, nivelando sus últimas líneas e introduciendo las designaciones más compactas para todos los vectores. En este caso obtendremos la siguiente tabla:

$$
\begin{array}{l}
e_1^{(t)}, \dots, e_{p_t}^{(t)}, \\
e_1^{(t-1)}, \dots, e_{p_t}^{(t-1)}, e_{p_t+1}^{(t-1)}, \dots, e_{p_{t-1}}^{(t-1)}, \\
\dots\dots\dots\dots\dots\dots \\
e_1^{(1)}, \dots, e_{p_t}^{(1)}, e_{p_t+1}^{(1)}, \dots, e_{p_{t-1}}^{(1)}, \dots, e_{p_2+1}^{(1)}, \dots, e_{p_1}^{(1)}.
\end{array}
\qquad (74.4)
$$

Los vectores que se hallen en la primera línea de esta tabla tienen la altura t, los vectores de la siguiente línea tienen la altura $t-1$, etc. Los vectores de la última línea tienen por altura 1, es decir, se transforman por el operador $A-\lambda_1E$ en un vector nulo. Cada columna de la tabla determina un subespacio invariante del operador $A-\lambda_1E$ y, por consiguiente, del operador A. Estos subespacios se llaman *cíclicos*. Los primeros p_1 subespacios cíclicos tienen dimensión t, los siguientes p_2-p_1 subespacios tienen una dimensión $t-1$, etc. Las últimas columnas determinan los subespacios cíclicos unidimensionales. *Todo el subespacio radical R_1 es la suma directa de los p_1 subespacios cíclicos radicales.*

Escribamos la matriz de un operador inducido en un subespacio cíclico. Supongamos, por ejemplo, que como base se han tomado los vectores $e_1^{(1)}, e_1^{(2)}, \ldots, e_1^{(t-1)}, e_1^{(t)}$. Puesto que

$$A(-\lambda_1E)e_1^{(1)}=0, \quad (A-\lambda_1E)e_1^{(2)}=e_1^{(1)}, \ldots, (A-\lambda_1E)e_1^{(t)}=e_1^{(t-1)},$$

tenemos

$$Ae_1^{(1)}=\lambda_1e_1^{(1)}, \; Ae_1^{(2)}=\lambda_1e_1^{(2)}+e_1^{(1)}, \ldots, Ae_1^{(t)}=\lambda_1e_1^{(t)}+e_1^{(t-1)}.$$

Por consiguiente, la matriz del operador inducido tendrá la forma

$$\begin{pmatrix} \lambda_1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_1 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & \lambda_1 & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_1 \end{pmatrix}.$$

Las matrices de tipo semejante se denominan *cajas canónicas de Jordan*.

Construyamos ahora la base de un espacio como reunión sucesiva de las bases de los subespacios radicales $R_1, R_2, \ldots, R_r$. Como base de cada subespacio radical R_i tomamos los vectores del tipo (74.4) ordenados por turno de abajo a arriba y de izquierda a derecha. La base del subespacio construida de la manera indicada se denomina *base radical*.

En una base radical la matriz J del operador A adquiere la así llamada *forma canónica de Jordan*. Es una matriz casi diagonal compuesta de las cajas de Jordan. Primeramente van dispuestas las cajas de Jordan que corresponden al valor propio λ_1, con la particularidad de que sus dimensiones, según las cuales están ordenadas, no crecen. Luego, en el mismo orden se disponen las cajas de Jordan

correspondientes a λ_r, etc. Así pues,

$$J = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & \lambda_1 \end{matrix}} & & & & 0 \\ & \ddots & & & \\ & & \boxed{\lambda_1} & & \\ & & & \ddots & \\ & & & & \boxed{\begin{matrix} \lambda_r & 1 & 0 \\ 0 & \lambda_r & 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & \lambda_r \end{matrix}} \\ & 0 & & & & \ddots \\ & & & & & & \boxed{\lambda_r} \end{pmatrix} \qquad (74.5)$$

Naturalmente, en el caso general algunas de las cajas de Jordan de órdenes inferiores pueden faltar.

La definición del operador en un espacio lineal determina la clase de matrices semejantes. El resultado obtenido significa que cualquier matriz cuadrada puede ser reducida, por medio de una transformación de semejanza, a la forma canónica de Jordan. Evidentemente, dos matrices cuadradas de un mismo orden son semejantes cuando, y sólo cuando, ambas tienen las formas de Jordan iguales. Por ello, con la base fijada:

Dos matrices cuadradas de un mismo orden determinan el mismo operador en un espacio complejo cuando, y sólo cuando, tienen iguales las formas de Jordan.

Ejercicios.

1. Sea x un vector radical de altura ν, correspondiente al valor propio λ_i del operador A. Demuéstrese que si λ_i es una raíz de multiplicidad p del polinomio $\varphi(z)$, el vector $u = \varphi(A)x$ será un vector radical de altura $\tau = \max(0, \nu - p)$ correspondiente al mismo valor propio λ_i. ¿Qué podrá decirse acerca del vector u, si λ_i no es la raíz del polinomio $\varphi(z)$?

2. Sea x un vector no nulo arbitrario y sea $\varphi(z)$ un polinomio de grado mínimo, para el cual $\varphi(A)x = 0$. Demuéstrese que $\varphi(z)$ es el divisor del polinomio característico del operador A.

3. Demuéstrese que toda matriz cuadrada puede ser reducida, salvo una permutación de las cajas de Jordan, a la forma canónica de Jordan de un tipo único.

4. Demuéstrese que si una matriz es semejante a la matriz J de (74.5), será también semejante a la matriz J'.

5. Demuéstrese que las matrices cuadradas A y A' son matrices de un mismo operador.

6. Sea J una matriz canónica de Jordan. ¿Qué forma tendrá la matriz J^p para los números enteros positivos p?

§ 75. Operador conjugado

Ahora pasamos a la investigación de los operadores lineales que actúan en un espacio unitario. Por supuesto, todos los resultados obtenidos anteriormente respecto de los operadores en un espacio complejo tienen lugar también en el caso dado. Por esta razón estudiaremos aquí sólo las propiedades adicionales de los operadores relacionadas con el concepto de ortogonalidad. En algunos casos consideraremos también unos operadores que actúan de un espacio unitario en otro espacio unitario. El papel principal en nuestras investigaciones lo desempeñará el así llamado operador conjugado.

Sean dados dos espacios unitarios X, Y. Un operador A^*, que actúa de Y en X, se llama *conjugado* del operador A, que actúa de X en Y, si para cualesquiera vectores $x \in X$, $y \in Y$ se verifica la igualdad

$$(Ax, y) = (x, A^*y). \tag{75.1}$$

TEOREMA 75.1. *Para todo operador lineal A existe un operador conjugado A^* que es, además, único.*

DEMOSTRACIÓN. Elijamos en X una base ortonormalizada $e_1, e_2, \ldots, e_m$. Recordemos que para todo vector $x \in X$ tiene lugar la descomposición

$$x = \sum_{k=1}^{m} (x, e_k) e_k. \tag{75.2}$$

Si el operador A^* existe, entonces, de acuerdo con esta fórmula, para cualquier vector $y \in Y$ tenemos

$$A^*y = \sum_{k=1}^{m} (A^*y, e_k) e_k$$

o bien, tomando en consideración (75.1),

$$A^*y = \sum_{k=1}^{m} \overline{(e_k, A^*y)}\, e_k = \sum_{k=1}^{m} \overline{(Ae_k, y)}\, e_k = \sum_{k=1}^{m} (y, Ae_k)\, e_k. \tag{75.3}$$

Esto precisamente significa que si el operador A^* existe, es el único.

Ahora, convengamos en considerar la igualdad (75.3) como definición del operador A^*. Es fácil comprobar que el operador A^*, construido de tal modo, es lineal. El operador satisface también la igualdad (75.1). En efecto, teniendo en cuenta el carácter ortonormalizado del sistema $e_1, e_2, \ldots, e_m$ y tomando en consideración

(75.2), (75.3), obtenemos para cualesquiera vectores $x \in X$, $y \in Y$

$$(Ax, y) = \left(A \sum_{k=1}^{m} (x, e_k) e_k, y\right) = \sum_{k=1}^{m} (x, e_k)(Ae_k, y),$$

$$(x, A^*y) = \left(\sum_{k=1}^{m} (x, e_k) e_k, \sum_{k=1}^{m} (y, Ae_k) e_k\right) =$$

$$= \sum_{k=1}^{m} (x, e_k)\overline{(y, Ae_k)} = \sum_{k=1}^{m} (x, e_k)(Ae_k, y).$$

El teorema queda demostrado.

El operador conjugado A^* está relacionado con el operador A por ciertas correlaciones. He aquí algunas de éstas:

$$\begin{aligned} &(A^*)^* = A, \\ &(A+B)^* = A^* + B^*, \\ &(\alpha A)^* = \bar{\alpha} A^*, \\ &(AB)^* = B^* A^*, \\ &(A^*)^{-1} = (A^{-1})^*. \end{aligned} \qquad (75.4)$$

La raya por encima de α significa aquí la conjugación compleja. Todas las correlaciones se demuestran siguiendo el mismo esquema. Por ello investiguemos detalladamente sólo las propiedades primera y segunda.

Consideraremos un operador arbitrario A y un operador A^*, conjugado de A. A su vez, para el operador A^*, el operador conjugado será $(A^*)^*$. Ahora, para cualesquiera $x \in X$, $y \in Y$ tenemos

$$(y, (A^*)^* x) = (A^*y, x) = \overline{(x, A^*y)} = \overline{(Ax, y)} = (y, Ax).$$

El primer miembro es igual al segundo para cualquier vector y. Por consiguiente, $(A^*)^* x = Ax$. Pero, como la igualdad dada es verídica para todo vector x esto es testimonio de que $(A^*)^* = A$.

Supongamos ahora que el operador A actúa en el espacio X y es regular. Demostremos, al principio, que el operador A^* también es regular. Sea $A^*y = 0$. Conforme a (75.3), proviene que

$$\sum_{k=1}^{m} (y, Ae_k) e_k = 0.$$

El sistema de vectores $e_1, \ldots, e_m$ constituye una base, razón por la cual

$$(y, Ae_k) = 0 \qquad (75.5)$$

para cualesquiera $k = 1, 2, \ldots, m$. El operador A es regular y, por lo tanto, toda base se trasforma por él en otra base. Pero, en

este caso el sistema de los vectores $Ae_1, \ldots, Ae_m$ será también una base y de (75.5) se deduce que $y = 0$. De este modo, el núcleo del operador A^* sólo contiene un vector nulo, es decir, este operador es regular.

Tomemos unos vectores arbitrarios $x, y \in X$. Existen los únicos vectores u, v, para los cuales

$$Au = x, \qquad A^*v = y.$$

Encontramos, luego,

$$(x, (A^{-1})^* y) = (A^{-1}x, y) = (u, A^*v) = (Au, v) = (x, (A^*)^{-1} y).$$

El primer miembro es igual al segundo para todo vector x. Por consiguiente, $(A^{-1})^* y = (A^*)^{-1} y$. Por ser y arbitrario, esto significa que $(A^{-1})^* = (A^*)^{-1}$.

Muchas propiedades conjuntas de los operadores A y A^* pueden establecerse en el proceso de investigación de las matrices de dichos operadores. Supongamos que en el espacio X se ha elegido una base ortonormalizada $e_1, e_2, \ldots, e_m$ y en el espacio Y, una base ortonormalizada $q_1, q_2, \ldots, q_n$. Si X coincide con Y, consideraremos coincidentes también sus bases. Supongamos que la matriz A_{qe} con los elementos a_{ij} corresponde al operador A. En este caso

$$Ae_j = \sum_{i=1}^{n} a_{ij}q_i.$$

Por tanto, de acuerdo con (75.2) concluimos que

$$a_{ij} = (Ae_j, q_i). \tag{75.6}$$

Supongamos, luego, que al operador A^* en las mismas bases le corresponda la matriz A^*_{eq} con los elementos a^*_{ij}. Conforme a la fórmula (75.6),

$$a^*_{ij} = (A^*q_j, e_i)$$

Comparando los elementos a_{ij}, a^*_{ij} y tomando en consideración (75.1), encontramos

$$a^*_{ij} = (A^*q_j, e_i) = \overline{(e_i, A^*q_j)} = \overline{(Ae_i, q_j)} = \bar{a}_{ji}.$$

Esta fórmula nos permite enunciar la siguiente definición.

La matriz A^* de dimensiones $m \times n$ con los elementos a^*_{ij} se denomina *conjugada* de la matriz A de dimensiones $n \times m$ con los elementos a_{ij}, siempre que $a^*_{ij} = \bar{a}_{ji}$ para cualesquiera i, j.

De este modo, en cualesquiera bases ortonormalizadas, a los operadores conjugados les corresponden unas matrices conjugadas. Las matrices conjugadas satisfacen, evidentemente, todas las correlaciones (75.4). La matriz conjugada A^* está ligada con la matriz A por las operaciones de transposición y conjugación compleja. A saber,

$$A^* = (\bar{A}') = (\bar{A})'. \tag{75.7}$$

Aquí la raya sobre la A y A' significa que todos los elementos de la matriz se sustituyen por los conjugados complejos.

El rango del operador coincide con el rango de su matriz. Por ello, de la fórmula (75.7) proviene que los operadores A y A^* tienen rangos iguales.

Designemos mediante $N \subset X$, $N^* \subset Y$ y $T \subset Y$, $T^* \subset X$ los núcleos y los campos de valores de los operadores A y A^*, respectivamente. Si $x \in N$, entonces $Ax = 0$ y $(x, A^*y) = 0$. Esto significa que el campo de valores del operador A^* es un subespacio ortogonal al núcleo del operador A. Naturalmente, el campo de valores del operador A es también ortogonal al núcleo del operador A^*. Por ser iguales las dimensiones de los subespacios T, T^* y las correlaciones (58.4), concluimos que

$$X = N \oplus T^*, \qquad Y = N^* \oplus T. \qquad (75.8)$$

La base $y_1, y_2, \ldots, y_m$ del espacio unitario X se denomina *dual* con relación a la base $x_1, x_2, \ldots, x_m$ del mismo espacio, si se verifica

$$(x_i, y_j) = \begin{cases} 0 & \text{para } i \neq j, \\ 1 & \text{para } i = j. \end{cases}$$

La base dual se usa con frecuencia para la investigación de las propiedades conjuntas de los operadores A y A^* que actúan en un mismo espacio. Demostremos, al principio, que toda base dispone de una base dual que es, además, única. Sea $x_1, x_2, \ldots, x_m$ una base arbitraria. Para cualquier j el vector y_j debe ser ortogonal a los vectores $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_m$, y, por lo tanto, ortogonal a la cápsula lineal L_j construida sobre estos vectores. De aquí se deduce que el vector y_j se halla en un subespacio unidimensional $L_j^{\perp}$. La condición de normalización $(x_j, y_j) = 1$ lo determina de un modo único.

Es obvio que una base será dual con relación a sí misma cuando, y sólo cuando, sea ortonormalizada. La relación de dualidad de las bases es simétrica, por lo cual resulta razonable hablar de un par de bases mutuamente duales. Las bases mutuamente duales se llaman *biortonormalizadas*.

TEOREMA 75.2. *Si el operador A tiene en una base la matriz J, entonces, en la base dual con relación a la dada, el operador conjugado A^* tiene la matriz J^*.*

DEMOSTRACIÓN. Supongamos que en la base ortonormalizada $e_1, e_2, \ldots, e_m$ los operadores A y A^* tienen las bases respectivas A_e y A_e^*, mientras que en la base $x_1, x_2, \ldots, x_m$ el operador A tiene la matriz J. Designemos con P la matriz de la transformación de coordenadas al pasar de la base $e_1, e_2, \ldots, e_m$ a la base $x_1, x_2, \ldots, x_m$. Entonces, de acuerdo con la fórmula (64.5), tenemos

$$J = P^{-1}A_eP.$$

Tomando una conjugación matricial de los miembros primero y segundo de esta igualdad, encontramos

$$J^* = P^* A_e^* (P^{-1})^*$$

o bien, lo que es igual,

$$J^* = ((P^{-1})^*)^{-1} A_e^* ((P^{-1})^*).$$

La correlación obtenida muestra que el operador conjugado A^* tiene la matriz J^* en la base $y_1, y_2, \ldots, y_m$, para la cual la matriz de la transformación de coordenadas, al pasar de la base $e_1, e_2, \ldots, e_m$, es $(P^{-1})^*$. Según la fórmula (63.3), las coordenadas de los vectores $z_1, z_2, \ldots, z_m$ en la base $e_1, e_2, \ldots, e_m$ son, en realidad, los elementos de las columnas de la matriz P; las coordenadas de los vectores $y_1, y_2, \ldots, y_m$ en la base $e_1, e_2, \ldots, e_m$ no son otra cosa que los elementos de las columnas de matriz $(P^{-1})^*$. El cálculo de los productos escalares de dos en dos de los vectores, pertenecientes a la base $z_1, z_2, \ldots, z_m$, y de los vectores, pertenecientes a la base $y_1, y_2, \ldots, y_m$, es equivalente al cálculo de los elementos de la matriz $P' \overline{(P^{-1})^*}$. Pero

$$P' \overline{(P^{-1})^*} = P' \overline{\overline{(P^{-1})'}} = P' (P^{-1})' = (P^{-1}P)' = E.$$

Por consiguiente, la base $y_1, y_2, \ldots, y_m$ es dual respecto a la base $z_1, z_2, \ldots, z_m$.

El teorema demostrado permite señalar muchos corolarios. Si, por ejemplo, J es una matriz canónica de Jordan, entonces en su diagonal se disponen los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_m$. Mas, los valores propios de la matriz J^* son $\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_m$. Por ello todos los valores propios del operador A^* son complejos conjugados respecto de los valores propios del operador A. Si el operador A es de estructura simple el teorema 75.2 permite afirmar que el operador conjugado A^* es también de estructura simple. En este caso los sistemas básicos de los vectores propios de los operadores A y A^* pueden elegirse de tal modo que sean biortonormalizados, etc.

Ejercicios.

1. Supongamos que las coordenadas de los vectores de cierta base de un espacio euclídeo, dadas en la base ortonormalizada $e_1, e_2, \ldots, e_m$, forman columnas de la matriz A. Demuéstrese que las coordenadas de los vectores de la base dual, dadas en la misma base $e_1, e_2, \ldots, e_m$, forman columnas de la matriz A^{-1}.

2. ¿De qué modo están ligados entre sí los polinomios característicos de los operadores A y A^*?

3. Demuéstrese que si un subespacio es invariante respecto del operador A, su complemento ortogonal es invariante respecto de A^*.

4. Demuéstrese que todo vector propio del operador A, correspondiente al valor propio λ, es ortogonal a cualquier vector propio del operador A^* correspondiente al valor propio $\mu \neq \bar{\lambda}$.

5. Demuéstrese que cualquier vector radical del operador A correspondiente al valor propio λ es ortogonal a todo vector radical del operador A^* correspondiente al valor propio $\mu \neq \lambda$.

§ 76. Operador normal

El hecho de que en un espacio existan una base ortonormalizada y una base formada por los vectores propios del operador lineal, es de mucha importancia para realizar las más diversas investigaciones. Por esto, nuestra tarea próxima consistirá en estudiar una clase de operadores que tienen en un espacio unitario sistemas básicos ortonormalizados compuestos de vectores propios. Operadores de tal índole existen a ciencia cierta. Por ejemplo, entre ellos figuran todos los operadores escalares.

TEOREMA 76.1. (*teorema de Schur*). *Para todo operador lineal en un espacio unitario existe una base ortonormalizada en la cual la matriz del operador es triangular.*

DEMOSTRACIÓN. Examinemos, por ejemplo, el caso de una matriz triangular derecha. De acuerdo con el teorema 72.1, para todo operador A existen los subespacios invariantes L_p, $p = 1, 2, \ldots, m$, tales que la dimensión de L_p es igual a p y cualquier subespacio de índice menor figura en todos los subespacios de índices mayores. La base buscada se construirá de la manera siguiente. A título de vector e_1 se tomará cualquier vector normalizado de L_1. A título de e_2 se tomará un vector normalizado de L_2, ortogonal al subespacio L_1, etc. A título de vector e_m se tomará un vector normalizado de L_m que sea ortogonal al subespacio L_{m-1}. La base $e_1, e_2, \ldots, e_m$ es ortonormalizada y, como se ha observado en el § 72, la matriz del operador en tal base es triangular derecha.

El operador lineal A se llama *normal*, si es permutable con su operador conjugado, es decir, si

$$AA^* = A^*A.$$

Probemos que los operadores normales, y sólo ellos, tienen en un espacio unitario sistemas básicos de vectores propios ortonormalizados.

Para el estudio de estos operadores resulta útil la siguiente observación. Si una matriz triangular es permutable con su matriz conjugada ésta es diagonal. En efecto, sea, por ejemplo, la matriz B de orden m triangular derecha y supongamos que $B^*B = BB^*$. Designemos mediante b_{ij} los elementos de la matriz B. La condición de que los elementos diagonales de la matriz $B^*B - BB^*$ son iguales a cero nos proporciona el siguiente sistema de ecuaciones res-

pecto de los elementos no diagonales de la matriz B:

$$\begin{aligned}
-|b_{12}|^2 - |b_{13}|^2 - |b_{14}|^2 - \ldots - |b_{1m}|^2 &= 0,\\
|b_{12}|^2 - |b_{23}|^2 - |b_{24}|^2 - \ldots - |b_{2m}|^2 &= 0,\\
|b_{13}|^2 + |b_{23}|^2 - |b_{34}|^2 - \ldots - |b_{3m}|^2 &= 0,\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots &\\
|b_{1m}|^2 + |b_{2m}|^2 + |b_{3m}|^2 + \ldots + |b_{m-1,m}|^2 &= 0.
\end{aligned}$$

La única solución de este sistema es una solución nula, por lo que se demuestra precisamente la validez de la observación enunciada.

TEOREMA 76.6 *Para que un operador en un espacio unitario sea normal, es necesario y suficiente que tenga un sistema básico de vectores propios ortonormalizados.*

DEMOSTRACIÓN Sea A un operador normal. Elijamos, de acuerdo con el teorema 76.4, una base ortonormalizada en la cual la matriz del operador sea triangular. En la misma base al operador A^* le corresponde una matriz triangular conjugada. Por hipótesis, el operador A es normal, por lo cual las matrices de los operadores A y A^* en la base elegida deben ser conmutables. Conforme a la observación enunciada anteriormente, estas matrices son diagonales. Así pues, hemos construido una base ortonormalizada en la cual la matriz del operador tiene una forma diagonal. Esto significa que la base dada está compuesta íntegramente de los vectores propios del operador.

Supongamos ahora que el operador A tiene un sistema básico de vectores propios ortonormalizados. En este caso la matriz del operador A en la base compuesta de los vectores citados será diagonal. Mas, en la misma base, al operador A^* le corresponde una matriz conjugada, la cual, evidentemente, será también diagonal. Las matrices diagonales son siempre conmutables y, por lo tanto, son conmutables también los operadores A y A^*.

En el transcurso de la demostración del teorema hemos probado que si el operador A es normal, entonces, en una base compuesta de vectores propios ortonormalizados no sólo la matriz del operador A, sino también la matriz del operador A^* será diagonal. Esto nos lleva al

COROLARIO. *Si el operador A es normal, todo sistema ortonormalizado de vectores propios del operador A es un sistema ortonormalizado de vectores propios del operador A^*, y viceversa.*

COROLARIO. *Si el operador A es normal, los valores propios de los operadores A y A^*, correspondientes al vector propio común, son complejos conjugados.*

Efectivamente, si $Ax = \lambda x$ y $A^*x = \mu x$, entonces, de acuerdo con (75.1), para todo vector propio normalizado x tendremos

$$\lambda = (\lambda x, x) = (Ax, x) = (x, A^*x) = (x, \mu x) = \bar{\mu}.$$

Este hecho es justo, por supuesto, para cualquier operador A que tiene unos vectores propios comunes con el operador A^*. El carácter normal del operador A garantiza *la presencia* de los vectores comunes.

La significación de los operadores normales en la teoría general se dicta por dos circunstancias. En primer lugar, esta clase de operadores es más simple en un espacio unitario. En segundo lugar, la investigación de un operador arbitrario normal se reduce frecuentemente a la investigación de los operadores normales.

Ejercicios.

1. Sea A un operador lineal arbitrario y α, β unos números complejos iguales en módulo. Demuéstrese que el operador $\alpha A + \beta A^*$ es normal.
2. Sea A un operador normal. Demuéstrese que para todo polinomio $\varphi(z)$ el operador $\varphi(A)$ será normal.
3. Demuéstrese que para un operador normal cualquier operador inducido será normal.
4. Demuéstrese que el operador A es normal cuando, y sólo cuando, para todo subespacio invariante L el complemento ortogonal $L^\perp$ es también invariante.
5. Sea A un operador de estructura simple en un espacio complejo. Demuéstrese que al prefijar de modo adecuado un producto escalar en un espacio, el operador A puede siempre hacerse normal.

§ 77. Operadores unitario y hermitiano

Entre los operadores normales son de mayor empleo los operadores de dos tipos: unitarios y hermitianos.

Un operador lineal U se llama *unitario*, si el operador conjugado U^* coincide con el inverso U^{-1}, es decir,

$$UU^* = U^*U = E.$$

TEOREMA 77.1. *Un operador normal U es unitario cuando, y sólo cuando, todos sus valores propios son iguales en módulo a la unidad.*

DEMOSTRACIÓN. Sea U un operador unitario. Elijamos cualquiera de sus valores propios λ y un vector propio normalizado x que corresponda a λ. Tenemos

$$1 = (x, x) = (x, U^*Ux) = (Ux, Ux) = (\lambda x, \lambda x) = \lambda\cdot\overline{\lambda}\,(x, x) = |\lambda|^2.$$

Supongamos ahora que todos los valores propios del operador normal U son iguales en módulo a la unidad. Designemos mediante $x_1, \ldots, x_m$ los vectores propios ortonormalizados del operador U y mediante $\lambda_1, \ldots, \lambda_m$, sus valores propios. Por hipótesis, $|\lambda_i| = 1$ para todo i. Recordemos que para el operador conjugado U^* los vectores $x_1, \ldots, x_m$ siguen siendo propios, pero corresponden a los valores propios $\overline{\lambda}_1, \ldots, \overline{\lambda}_m$. Tomemos un vector arbitrario x y des-

compongámoslo según los vectores propios del operador U

$$x = \alpha_1 x_1 + \ldots + \alpha_m x_m.$$

Ahora calculamos

$$U^*Ux = U^*(Ux) = U^*(\alpha_1\lambda_1 x_1 + \ldots + \alpha_m\lambda_m x_m) =$$
$$= \alpha_1\lambda_1\bar{\lambda}_1 x_1 + \ldots + \alpha_m\lambda_m\bar{\lambda}_m x_m = \alpha_1 x_1 + \ldots + \alpha_m x_m = x.$$

Puesto que x es un vector arbitrario, esto significa que $U^*U = E$. De modo análogo se demuestra que $UU^* = E$.

TEOREMA 77.2. *El operador U es unitario cuando, y sólo cuando, para cualesquiera dos vectores el producto escalar de éstos es igual al producto escalar de sus imágenes.*

DEMOSTRACION. Sea U un operador unitario, entonces para cualesquiera dos vectores x, y tenemos

$$(x, y) = (x, U^*Uy) = (Ux, Uy). \qquad (77.1)$$

Supongamos ahora que para cierto operador U se verifican las igualdades (77.1), cualesquiera que sean los vectores x, y. De aquí se deduce que

$$(x, (U^*U - E)\, y) = 0.$$

Puesto que los vectores x, y son arbitrarios, esto significa que $U^*U = E$. El operador U es regular, dado que en el caso contrario la igualdad $U^*U = E$ sería imposible. Por consiguiente, el operador U^{-1} existe. Al multiplicar la igualdad $U^*U = E$ a la izquierda por el operador U y a la derecha, por el operador U^{-1}, obtenemos otra igualdad: $UU^* = E$. De este modo, el operador U es unitario.

COROLARIO. *El operador U es unitario cuando, y sólo cuando, o bien $UU^* = E$ o bien $U^*U = E$.*

COROLARIO. *Todo operador unitario transforma cualquier sistema ortonormalizado de vectores en otro sistema, también ortonormalizado.*

COROLARIO. *Si el operador lineal U transforma una base ortonormalizada en otra base ortonormalizada, entonces U es un operador unitario.*

En efecto, sea $x_1, \ldots, x_m$ una base ortonormalizada, $Ux_i = y_i$ e $y_1, \ldots, y_m$, también una base ortonormalizada. Tomemos dos vectores arbitrarios x, y. Si

$$x = \sum_{i=1}^{m} \alpha_i x_i, \qquad y = \sum_{i=1}^{m} \beta_i x_i,$$

entonces

$$(x, y) = \sum_{i=1}^{m} \alpha_i \bar{\beta}_i.$$

Por ser lineal el operador U, tenemos

$$Ux = \sum_{i=1}^{m} \alpha_i y_i, \qquad Uy = \sum_{i=1}^{m} \beta_i y_i.$$

Por eso, tenemos nuevamente

$$(Ux, Uy)=\sum_{i=1}^{n}\alpha_i\bar{\beta}_i.$$

Así pues, las igualdades (77.1) son válidas para cualesquiera vectores x, y.

Hemos de notar que podríamos definir el operador unitario como operador *isométrico*, es decir, un operador que conserva invariables las longitudes de todos los vectores. Esto proviene del teorema 77.2 y de la siguiente correlación, fácilmente comprobable:

$$(x, y)=\frac{|x+y|^2-|x-y|^2+i|x+iy|^2-i|x-iy|^2}{4}.$$

Un operador lineal H se denomina *hermitiano* o *autoconjugado*, al coincidir con su operador conjugado, es decir, si

$$H=H^*.$$

TEOREMA 77.3. *Un operador normal H es hermitiano cuando, y sólo cuando, todos sus valores propios son números reales.*

DEMOSTRACIÓN. Sea H un operador hermitiano. Tomemos cualquier valor propio λ de este operador y un vector propio normalizado x que corresponde a λ. Tenemos

$$\lambda=(\lambda x, x)=(Hx, x)=(x, H^*x)=(x, Hx)=(x, \lambda x)=\bar{\lambda},$$

es decir, λ es un número real. Supongamos ahora que el operador normal H tiene valores propios reales. Entonces, en la base compuesta de los vectores propios ortonormalizados del operador H las matrices de los operadores H y H^* coincidirán. Por consiguiente, son también coincidentes los operadores, es decir, H es un operador hermitiano.

El operador hermitiano H se llama *no negativo (definido positivo)* si para todo vector (no nulo) x se verifica la desigualdad

$$(Hx, x)\geqslant 0 \ (>0).$$

TEOREMA 77.4. *El operador hermitiano H es no negativo (definido positivo) cuando, y sólo cuando, todos sus valores propios son no negativos (positivos).*

DEMOSTRACIÓN. Elijamos una base ortonormalizada compuesta de los vectores propios $x_1, \ldots, x_m$ del operador hermitiano H. En este caso, de la descomposición

$$x=\xi_1x_1+\ldots+\xi_mx_m$$

se desprende, para un vector arbitrario x, que

$$(Hx, x)=\lambda_1|\xi_1|^2+\ldots+\lambda_m|\xi_m|^2.$$

De aquí se deduce que si todos los valores propios del operador hermitiano son no negativos (positivos), entonces el mismo operador también es no negativo (definido positivo). Al hacer $x=x_1$, obtene-

mos

$$(Hx_i,\ x_i) = \lambda_i$$

para todo i. Por ello, todos los valores propios de un operador no negativo (definido positivo) son no negativos (positivos).

De lo dicho se infiere que un operador definido positivo es regular no negativo. Entre todos los operadores hermitianos los operadores no negativos y definidos positivos desempeñan un papel de especial importancia. He aquí algunas de sus propiedades.

Si H y S son unos operadores definidos positivos, el operador $\alpha H +$ $+ \beta S$ será definido positivo para cualesquiera números no negativos α, β, no iguales a cero simultáneamente.

Efectivamente, el operador $\alpha H + \beta S$ es hermitiano, cualesquiera que sean los números reales α, β. Si estos números son negativos y no iguales a cero a la vez, entonces

$$((\alpha H + \beta S)\, x,\ x) = \alpha\,(Hx,\ x) + \beta\,(Sx,\ x) > 0$$

cuando $x \neq 0$.

Si el operador H es definido positivo, el operador H^{-1} será también definido positivo.

En efecto, como $H = H^*$, entonces $H^{-1} = (H^*)^{-1} = (H^{-1})^*$, es decir, el operador H^{-1} es hermitiano. Los valores propios del operador H^{-1} son magnitudes inversas respecto a los valores propios del operador H. Por ello son positivos y el operador H^{-1} es definido positivo.

Si H es definido positivo y A es un operador regular arbitrario, entonces los operadores A^*HA y AHA^* son definidos positivos.

Es fácil comprobar que estos operadores son hermitianos. En virtud de que el operador A es regular para cualquier $x \neq 0$, tendremos $Ax \neq 0$ y $A^*x \neq 0$. Por esta razón

$$(A^*HAx,\ x) = (HAx,\ Ax) > 0,\quad (AHA^*x,\ x) = (HA^*x, A^*x) > 0$$

para $x \neq 0$. De aquí se deduce, en particular, que para todo operador regular A los operadores A^*A y AA^* son definidos positivos. Si A es un operador degenerado, los operadores A^*A y AA^* son no negativos.

Para todo operador no negativo H existe un operador no negativo S tal que $S^2 = H$.

En efecto, sean $\lambda_1, \ldots, \lambda_m$ los valores propios del operador H y $x_1, \ldots, x_m$ los vectores propios ortonormalizados correspondientes. En este caso $Hx_i = \lambda_i x_i$ para todo i. Definamos el operador S mediante las igualdades $Sx_i = \sqrt{\lambda_i}x_i$. El operador S es no negativo, puesto que tiene un sistema básico de vectores propios ortonormalizados $x_1, \ldots, x_m$ que corresponden a los valores propios no negativos $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}$. Además, $S^2x_i = Hx_i = \lambda_i x_i$. De este modo, los operadores S^2 y H coinciden sobre los vectores de la base $x_1, \ldots$

$\ldots, x_m$, a consecuencia de lo cual ellos coinciden también sobre todos los vectores, es decir, $S^2 = H$.

El operador no negativo S se denomina raíz *cuadrada aritmética* de un operador no negativo H, si $S^2 = H$.

Merece subrayar que todos los vectores propios de los operadores S y H coinciden. Efectivamente, supongamos que $\lambda_1, \ldots, \lambda_r$ y $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}$ son todos valores propios distintos de los operadores H y S, respectivamente. Designemos mediante X_i (Y_i), $i = 1, 2, \ldots, r$, al subespacio propio del operador H (S) que contiene todos los vectores propios correspondientes al valor propio λ_i $(\sqrt{\lambda_i})$. Las sumas directas de los subespacios propios $X_1, \ldots, X_r$ e $Y_1, \ldots, Y_r$ coinciden con todo el espacio. Por esto

$$\dim X_1 + \ldots + \dim X_r = \dim Y_1 + \ldots + \dim Y_r. \qquad (77.2)$$

Está claro que para todo i se tiene $Y_i \subset X_i$, es decir, $\dim Y_i \leqslant \dim X_i$. Por consiguiente, la igualdad (77.2) puede verificarse sólo cuando para cualquier i se cumple la igualdad $\dim Y_i = \dim X_i$, es decir, cuando $Y_i = X_i$.

Así pues, los valores propios y los vectores propios del operador S se determinan unívocamente por el operador H. Puesto que S es un operador hermitiano, la raíz aritmética del operador H puede ser sólo *única*.

Ejercicios.

1. Demuéstrese que el conjunto de todos los operadores unitarios en un espacio unitario dado forma un grupo de multiplicación.

2. Demuéstrese que el conjunto de todos los operadores hermitianos en un espacio unitario dado forma un grupo de adición.

3. Supongamos que el operador A es hermitiano y el operador B es definido positivo. Demuéstrese que los valores propios de los operadores BA y $B^{-1}A$ son reales.

4. Demuéstrese que si A, B son unos operadores definidos positivos, todos los valores propios del operador BA son positivos.

5. Demuéstrese que si A, B son unos operadores definidos positivos conmutables, el operador BA será también definido positivo.

6. Demuéstrese que si A es un operador definido positivo en un espacio unitario, la función $(x, y)_A = (Ax, y)$ satisface todos los axiomas del producto escalar.

§ 78. Operadores A^*A y AA^*

Si el operador A actúa de un espacio unitario X en otro espacio unitario Y, entonces en X queda definido el operador A^*A y en Y, el operador AA^*. En lo que sigue estos operadores desempeñarán un papel considerable, por lo cual nos dedicaremos, ahora, a su estudio.

De las propiedades primera y cuarta (75.4) se deduce que los operadores A^*A y AA^* son hermitianos. Más aún, son no negativos, por-

que para cualesquiera vectores $x \in X$, $y \in Y$ tenemos

$$(A^*Ax, x) = (Ax, Ax) \geqslant 0$$

$$(AA^*y, y) = (A^*y, A^*y) \geqslant 0.$$

Por eso en el espacio X existe un operador no negativo G y en el espacio Y, un operador no negativo F tales que

$$A^*A = G^2, \qquad AA^* = F^2.$$

Los operadores G y F que satisfacen estas correlaciones son únicos.

Cualquiera que sea el operador A, el operador A^*A tiene un sistema *ortonormalizado* de los vectores propios $x_1, x_2, \ldots, x_m$. Este sistema se transforma siempre por el operador A en un sistema *ortogonal*. Efectivamente, sea

$$A^*Ax_k = \rho_k^2 x_k, \qquad \rho_k \geqslant 0 \tag{78.1}$$

para todo $k = 1, 2, \ldots, m$. Entonces

$$(Ax_k, Ax_l) = (A^*Ax_k, x_l) = \rho_k^2 (x_k, x_l) = 0$$

para $k \neq l$. Además, para todo k

$$|Ax_k| = \rho_k,$$

por lo cual el vector Ax_k es distinto del vector nulo cuando, y sólo cuando, el valor propio ρ_k^2 del operador A^*A no es nulo.

El vector *no nulo* Ax_k es un vector propio del operador AA^* y corresponde al valor propio ρ_k^2. En efecto, de acuerdo con (78.1)

$$AA^*(Ax_k) = A(A^*Ax_k) = A(\rho_k^2 x_k) = \rho_k^2 Ax_k.$$

De este modo, todos los valores propios no nulos del operador A^*A son valores propios del operador AA^*. Será cierta también, por supuesto, la afirmación inversa. Por esta razón los valores propios *no nulos* de los operadores A^*A y AA^* siempre *coinciden*.

Los valores propios de los operadores A^*A y AA^* se designarán mediante $\rho_1^2, \rho_2^2, \ldots$ En este caso puede considerarse, sin limitar la generalidad del razonamiento

$$\rho_1^2 \geqslant \rho_2^2 \geqslant \ldots \geqslant \rho_r^2 > 0,$$

mientras que los demás valores ρ_k^2 son iguales a cero. Es evidente que los valores propios de los operadores A^*A y AA^* se diferencian sólo en la multiplicidad del valor propio nulo. La multiplicidad del operador A^*A es $(m - r)$ y la del operador AA^*, $(n - r)$.

Se llaman *números singulares* (*principales*) *del operador* A los valores aritméticos de las raíces cuadradas de los valores propios comunes de los operadores A^*A y AA^*.

Haciendo uso de los vectores propios de los operadores A^*A y AA^*, se pueden construir en los espacios X e Y unas bases ortonormalizadas con ayuda de las cuales se describe y se investiga con facili-

dad la acción de los operadores A y A^*. Tomemos por base en el espacio X un sistema ortonormalizado $x_1, \ldots, x_m$ de vectores propios del operador A^*A. Según se deduce de (75.8), los vectores $x_1, \ldots, x_r$ forman una base en T^*, mientras que los vectores $x_{r+1}, \ldots, x_m$ forman una base en N. La base ortonormalizada $y_1, \ldots, y_n$ en el espacio Y se construirá de la manera siguiente. A título de $y_1, \ldots, y_r$ se tomarán los vectores obtenidos después de normalizar $Ax_1, \ldots, Ax_r$. Estos vectores forman una base en T. Por $y_{r+1}, \ldots, y_n$ se tomará cualquier base ortonormalizada en N^*. Está claro que los vectores $y_1, \ldots, y_n$ son propios para el operador AA^* y forman una base en Y. Tomando en consideración que $|Ax_k| = \rho_k$, deducimos que

$$Ax_k = \begin{cases} \rho_k y_k, & k \leqslant r, \\ 0, & k > r. \end{cases} \tag{78.2}$$

Al multiplicar estas igualdades por el operador A^* y teniendo en cuenta (78.1), obtenemos

$$A^*y_k = \begin{cases} \rho_k x_k, & k \leqslant r, \\ 0, & k > r. \end{cases} \tag{78.3}$$

Las bases ortonormalizadas en los espacios X, Y, ligadas con los operadores A, A^* mediante las correlaciones (78.2), (78.3), llevan el nombre de *bases singulares*.

Si los espacios X, Y son diferentes, en las bases singulares puede escribirse la matriz del operador A. Designémosla con Λ. Conforme a (78.2) esta matriz tiene la forma

$$\Lambda = \begin{pmatrix} \rho_1 & & & & & \\ & \rho_2 & & 0 & & \\ & & \ddots & & & \\ & & & \rho_r & & \\ & 0 & & & 0 & \\ & & & & & \ddots \end{pmatrix} \tag{78.4}$$

Si los espacios X, Y coinciden, las bases singulares, por regla general, no se usan para la obtención de la matriz del operador. Sin embargo, las correlaciones (78.2), (78.3) quedan en vigor.

Ejercicios.

1. Demuéstrese que los núcleos de los operadores A, A^*A (A^*, AA^*) coinciden, como también son coincidentes los campos de valores de los operadores A, AA^* (A^*, A^*A).
2. Demuéstrese que si $\dim X > \dim Y$ ($\dim X < \dim Y$), el operador A^*A (AA^*) es degenerado.
3. Demuéstrese que los números singulares no varían cuando el operador A se multiplica por cualesquiera operadores unitarios.

4. Supongamos que el operador A actúa en el espacio X y que todos sus números singulares son distintos dos a dos. Demuéstrese que las bases singulares se determinan unívocamente, salvo la multiplicación de cada uno de los vectores por un número que en módulo es igual a la unidad.

5. Demuéstrese que los números singulares de un operador normal coinciden con los módulos de los valores propios.

6. Demuéstrese que los números singulares del operador A^{-1} son inversos a los números singulares del operador A, mientras que las bases singulares de ambos operadores coinciden.

7. Supongamos que el operador A actúa en un espacio unitario m-dimensional X. Designemos mediante $\lambda_1, \ldots, \lambda_m$ sus valores propios y mediante $\rho_1, \ldots, \rho_m$ los números singulares. Demuéstrese que

$$\sum_{k=1}^{m} |\lambda_k|^2 \leqslant \sum_{k=1}^{m} \rho_k^2, \qquad \prod_{k=1}^{m} |\lambda_k| = \prod_{k=1}^{m} \rho_k.$$

8. Demuéstrese que si $|\lambda_k| = \rho_k$ para todo $k = 1, 2, \ldots, m$, entonces el operador es normal.

§ 79. Descomposiciones de un operador arbitrario

Una de las circunstancias que determina la significación de los operadores unitario y hermitiano consiste en la posibilidad de representar, sirviéndose de estos operadores, un operador lineal arbitrario.

Supongamos que un operador lineal arbitrario A actúa en el espacio unitario X. Mostremos que dicho operador siempre puede ser representado en la forma

$$A = H_1 + iH_2, \tag{79.1}$$

donde H_1 y H_2 son unos operadores hermitianos. Efectivamente, si esta descomposición existe, entonces

$$A^* = H_1 - iH_2.$$

Pero, en este caso

$$H_1 = \frac{1}{2}(A + A^*), \qquad H_2 = \frac{1}{2i}(A - A^*).$$

Las fórmulas obtenidas determinan precisamente la descomposición (79.1). Puesto que

$$H_1H_2 - H_2H_1 = \frac{1}{2i}(A^*A - AA^*),$$

entonces del hecho de que el operador A es normal proviene la conmutatividad de los operadores H_1, H_2, y viceversa.

Sea $x_1, \ldots, x_m$ un sistema ortonormalizado de vectores propios del operador A^*A. De acuerdo con (78.2), existe un sistema ortonormalizado $y_1, \ldots, y_m$ de vectores propios del operador AA^* tal que para todo k se verifica

$$Ax_k = \rho_k y_k. \tag{79.2}$$

Definamos ahora los operadores lineales F y U en el espacio X mediante las siguientes igualdades en los sistemas básicos de vectores:

$$Ux_k = y_k, \qquad Fy_k = \rho_k y_k. \tag{79.3}$$

Las correlaciones (79.2), (79.3) significan que se ha obtenido la descomposición

$$A = FU. \tag{79.4}$$

Aquí F es un operador hermitiano no negativo, puesto que tiene el sistema básico ortonormalizado de vectores propios $y_1, y_2, \ldots, y_m$ y los valores propios no negativos $\rho_1, \rho_2, \ldots, \rho_m$. El operador U es unitario, puesto que transforma el sistema ortonormalizado de vectores $x_1, x_2, \ldots, x_m$ en otro sistema ortonormalizado $y_1, y_2, \ldots, y_m$. Ha de ser notado que de (79.4) se infiere

$$AA^* = F^2, \tag{79.5}$$

es decir, F es la raíz cuadrada aritmética del operador AA^*.

La descomposición (79.4) se llama *descomposición polar* del operador A. Como que la raíz aritmética es única, el operador F en la descomposición polar será siempre único. El operador U será único sólo en el caso en que el operador A sea regular. En este caso $U = F^{-1}A$.

Otra vez observamos la relación directa existente entre el carácter normal del operador A y la conmutatividad de las componentes de una descomposición polar. Efectivamente, sea $UF = FU$ para cierto operador A, entonces

$$A^*A = U^*F^*FU = F^*U^*UF = F^2,$$

lo que, junto con (79.5), es testimonio de que el operador A es normal.

Supongamos ahora que el operador A es normal, es decir, $A^*A = AA^*$. De acuerdo con (79.4) tenemos $A = FU$. Por consiguiente, $A^* = U^*F$. La condición de que el operador es normal conduce a la igualdad $U^*F^2U = F^2$ o bien

$$F^2U = UF^2.$$

Tomando en consideración la segunda de las correlaciones (79.3), obtenemos

$$F^2(Uy_k) = \rho_k^2(Uy_k)$$

para todo $k = 1, 2, \ldots, m$, es decir, los vectores Uy_k son propios para el operador F^2. Como se ha observado anteriormente, los operadores F^2 y F tienen los mismos vectores propios, por lo cual

$$(FU)\,y_k = F(Uy_k) = \rho_k(Uy_k)$$

para todo $k = 1, 2, \ldots, m$. Por otra parte, conforme a la segunda de las correlaciones (79.3),

$$(UF)\,y_k = U\,(Fy_k) = U\,(\rho_k y_k) = \rho_k\,(Uy_k).$$

Las igualdades obtenidas muestran que los operadores FU y UF coinciden en el sistema básico de vectores $y_1, y_2, \ldots, y_m$. Por consiguiente, $UF = FU$.

Ejercicios.

1. Demuéstrese que si un operador es normal, los valores propios del operador H_1 (H_2) de (79.1) son las partes reales (imaginarias) de los valores propios del operador A.
2. Demuéstrese que si el operador A es normal, los valores propios del operador F (argumentos de los valores propios del operador U) de (79.4) son los módulos de los valores propios (argumentos de los valores propios no nulos) del operador A.
3. Demuéstrese que si el operador A es normal, entonces ambos operadores en la descomposición (79.1) tienen los mismos vectores propios que el operador A. ¿Qué puede decirse acerca de los vectores propios de los componentes de la descomposición (79.4)?

§ 80. Operadores en el espacio real

Al investigar los operadores lineales que actúan en un espacio real nos encontramos con algunas dificultades adicionales. Están relacionadas principalmente con el hecho de que no todo operador lineal tiene en el espacio real siquiera un solo vector propio.

Naturalmente, si el polinomio característico de un operador en un espacio real tiene sólo raíces reales, subsiste la analogía completa en la teoría. Varía, de hecho, sólo la terminología. A saber, las palabras «complejo, unitario, hermítico» se sustituyen respectivamente por las palabras «real, ortogonal, simétrico». Si, en cambio, el polinomio característico tiene, además, unas raíces complejas, la investigación de tal operador se hace más complicada.

Sea dado un espacio real $\mathbf{R}$. Consideraremos un conjunto de toda clase de pares $(x; y)$ de vectores x, y, de $\mathbf{R}$. Definamos las operaciones sobre dichos pares. Convengamos en considerar que

$$(x;\ y) + (u;\ v) = (x + u;\ y + v)$$

para cualesquiera dos pares, y para el número complejo $\xi + i\eta$ y el par $(x; y)$

$$(\xi + i\eta)\,(x;\ y) = (\xi x - \eta y;\ \eta x + \xi y)$$

Es fácil comprobar que el conjunto de todos los pares de vectores pertenecientes a $\mathbf{R}$, después de realizadas las operaciones de la manera indicada, representa en sí un espacio complejo $\mathbf{C}$.

El espacio construido C es de la misma dimensión que el espacio **R**. En efecto, sea $e_1, e_2, \ldots, e_m$ una base en **R**. Para todo par de vectores u, v de **R** tenemos

$$\left.\begin{aligned} u &= \alpha_1 e_1 + \ldots + \alpha_m e_m \\ v &= \beta_1 e_1 + \ldots + \beta_m e_m \end{aligned}\right\} \qquad (80.1)$$

donde α_i, β_i son los números reales. Pero de aquí proviene que

$$(u; v) = \sum_{k=1}^{m} (\alpha_k + i\beta_k)(e_k; 0). \qquad (80.2)$$

El sistema $(e_1; 0), \ldots, (e_m; 0)$ es linealmente independiente, por lo cual la dimensión del espacio C es igual a m.

Para toda base $e_1, \ldots, e_m$ en el espacio **R** y cualesquiera números reales $\alpha_1, \ldots, \alpha_m$ se cumple la igualdad

$$\sum_{k=1}^{m} (\alpha_k + i0)(e_k; 0) = \left(\sum_{k=1}^{m} \alpha_k e_k; 0\right).$$

Por consiguiente, entre todos los vectores u de **R** y todos los pares del tipo $(u; 0)$ de C existe una correspondencia biunívoca. Más aún, esta correspondencia es un isomorfismo, si nos limitamos a las operaciones con números reales.

Si todos los pares del tipo $(u; 0)$ se identifican con los propios vectores u de **R**, entonces de (80.1), (80.2) se desprende que el espacio C puede considerarse como un conjunto de elementos

$$w = u + iv,$$

donde $u, v \in$ **R**. En este caso se debe recordar, por supuesto, que en realidad los elementos u, v son los pares $(u; 0)$, $(v; 0)$ y la multiplicación por el número i y la adición se realizan conforme a las definiciones introducidas anteriormente. Cuando $v = 0$, obtenemos los elementos del espacio **R**. Es natural considerar este espacio como cierto conjunto de C. Los elementos del tipo $u + i0$ se llamarán *reales*, mientras que los elementos del tipo $u + iv$ y $u - iv$ se denominan *complejos conjugados*.

El espacio C se llama *extensión compleja del espacio real* **R**.

En la resolución de los diversos problemas en un espacio euclídeo podemos extender dicho espacio, de modo análogo, hasta obtener un espacio unitario. Examinemos la extensión compleja C del espacio euclídeo **R**. Para cualesquiera dos vectores

$$z = x + iy, \qquad w = u + iv$$

de C, convengamos en considerar, por definición, que

$$(z, w) = ((x, u) + (y, v)) + i((y, u) - (x, v)).$$

No es difícil establecer que el espacio C dotado de tal producto escalar se hace unitario. Además, el producto escalar para cualesquiera dos vectores de R se conserva.

Supongamos que el operador A actúa en el espacio R. Construyamos un operador nuevo $\tilde{A}$ que actúa en el espacio C y coincide con el operador A sobre los vectores de R. Para esto hagamos

$$\tilde{A}(u+iv)=Au+iAv.$$

Está claro que el operador $\tilde{A}$ es lineal y $\tilde{A}u=Au$ para todos los vectores $u \in R$.

El operador $\tilde{A}$ lleva el nombre de *extensión del operador A en el espacio complejo C*.

Ahora, en vez de estudiar el operador A en el espacio real R podemos considerar el operador $\tilde{A}$ en su espacio complejo C e investigar su operación en R, considerando este último como conjunto del espacio C. Esto es un procedimiento más usado, si para alguna situación en el espacio complejo no existe analogía correspondiente en el espacio real.

Supongamos que en el espacio complejo C se ha elegido una base real. En esta base la matriz del operador dilatado $\tilde{A}$ será real y coincidirá con la matriz del operador A en la misma base. De aquí se infiere que el polinomio característico del operador $\tilde{A}$ coincide con el polinomio característico del operador A y, por consiguiente, tiene coeficientes reales. Es evidente que

Si el polinomio característico del operador A que actúa en un espacio real R tiene raíz real, este último es un valor propio del operador $\tilde{A}$ y le corresponde al menos un vector propio real.

Consideraremos ahora una raíz compleja λ del polinomio característico del operador A. Es un valor propio del operador $\tilde{A}$ y le corresponde cierto vector propio w. Dado que los coeficientes del polinomio característico del operador $\tilde{A}$ son reales, el operador citado tendrá también un valor propio conjugado complejo $\bar{\lambda}$. El operador $\tilde{A}$ transforma vectores conjugados complejos en vectores conjugados complejos, por lo cual de $\tilde{A}w=\lambda w$ proviene $\tilde{A}\bar{w}=\bar{\lambda}\bar{w}$. Por consiguiente, a los valores propios conjugados complejos del operador $\tilde{A}$ les corresponden unos vectores conjugados complejos.

Si $\lambda \neq \bar{\lambda}$, entonces los vectores w, $\bar{w}$ serán linealmente independientes como vectores propios correspondientes a los diferentes valores propios.

Consideraremos los vectores x, y que se determinan del modo siguiente en términos de w, $\bar{w}$:

$$x=\frac{1}{2}(w+\bar{w}), \qquad y=\frac{1}{2i}(w-\bar{w}). \qquad (80.3)$$

Es fácil comprobar que dichos vectores son reales. Además, no es difícil establecer que si $\lambda = \mu + i\nu$, entonces

$$Ax = \mu x - \nu y, \qquad Ay = \nu x + \mu y.$$

Por ello, una cápsula lineal construida sobre los vectores (80.3) en el espacio R es un subespacio invariante del operador A. La matriz del operador inducido sobre este subespacio en la base (80.3) es la siguiente:

$$\begin{pmatrix} \mu & \nu \\ -\nu & \mu \end{pmatrix}.$$

Por consiguiente, el polinomio característico del operador inducido es igual a $(z - \mu)^2 + \nu^2$ o bien, lo que es igual, $z^2 - (\lambda + \bar{\lambda}) z + \lambda\bar{\lambda}$. Observemos que en el subespacio invariante construido el operador A no tiene ningún vector propio cuando $\nu \neq 0$. De este modo, hemos obtenido una deducción importante. A saber:

Si un polinomio característico del operador A, que actúa en el espacio real R, tiene una raíz compleja (no real), a esta raíz le corresponde en el espacio R un subespacio invariante bidimensional del operador A que no contiene vectores propios.

Esta deducción juega el mismo papel para investigar los operadores en un espacio real que desempeña la existencia por lo menos de un único vector propio para la investigación de los operadores en un espacio complejo. Eligiendo de un modo adecuado las bases en el espacio R, se puede reducir la matriz del operador a una forma semejante, en cierto sentido, o bien a la diagonal, o bien a la triangular, o bien a la forma canónica de Jordan. Tal procedimiento de investigación del operador es de uso relativamente raro, puesto que las formas canónicas reales están privadas de muchas ventajas que poseen las formas canónicas complejas. Es mucho más fácil y fructífero la investigación de la dilatación de un operador en un espacio complejo.

Ejercicios.

1. Demuéstrese que el campo de valores (el núcleo) del operador $\tilde{A}$ es la dilatación compleja del campo de valores (del núcleo) del operador A.

2. Supongamos que el operador dilatado $\tilde{A}$ es de estructura simple. Demuéstrese que en el espacio R se puede elegir tal base en la que la matriz del operador A sea de forma casi diagonal con las matrices en la diagonal de primero y segundo órdenes.

3. Demuéstrese que en el espacio real R de dimensión m todo operador tiene un subespacio invariante de dimensión $m - 1$ ó $m - 2$.

4. ¿Qué análogo tiene en un espacio real el teorema 72.1?

5. Demuéstrese que todo operador lineal que actúa en un espacio real de dimensión impar tiene por lo menos un vector propio.

§ 81. Matrices de tipo especial

Hemos considerado algunos operadores de tipo especial. Será natural suponer que las matrices de dichos operadores deben poseer ciertos rasgos específicos.

Una matriz compleja cuadrada U se llama *unitaria*, si la matriz conjugada U^* coincide con la inversa U^{-1}, es decir, si

$$UU^* = U^*U = E.$$

Recordemos que en una base ortonormalizada al operador conjugado le corresponde una matriz conjugada. Por consiguiente, la matriz de un operador unitario en la base ortonormalizada es unitaria.

Sean dadas en un espacio unitario dos bases ortonormalizadas cualesquiera. Construyamos la matriz de la transformación de coordenadas al pasar de una de estas bases a la otra. De acuerdo con la fórmula (63.3), las columnas de la matriz están compuestas por las coordenadas que tienen los vectores de la segunda base respecto de la primera. Mas la misma forma tiene también la matriz del operador lineal que transforma los vectores de la primera base en los de la segunda. Conforme al segundo corolario del teorema 77.2, este operador es unitario. Por esto

La matriz de la transformación de coordenadas al pasar de una base ortonormalizada a otra base ortonormalizada es unitaria.

Llamaremos dos matrices *semejantes unitarias*, si son semejantes y la matriz de la transformación de semejanza es unitaria. De las propiedades del operador unitario se deduce que toda matriz unitaria es semejante unitaria respecto de una matriz diagonal de elementos diagonales que, en módulo, son iguales a la unidad.

Se escriben con facilidad las correlaciones que definen los elementos de la matriz unitaria. Supongamos que la matriz U es de orden m. Designemos mediante u_{ij} sus elementos. Entonces, de la igualdad $UU^* = E$ se infiere que

$$\sum_{k=1}^{m} u_{ik}\bar{u}_{jk} = \begin{cases} 0, & \text{si } i \neq j, \\ 1, & \text{si } i = j. \end{cases}$$

Análogamente, de la igualdad $U^*U = E$ obtenemos:

$$\sum_{k=1}^{m} u_{ki}\bar{u}_{kj} = \begin{cases} 0, & \text{si } i \neq j, \\ 1, & \text{si } i = j. \end{cases}$$

De este modo, los sistemas de vectores columna y vectores fila de cualquier matriz unitaria representan en sí sistemas ortonormalizados.

La matriz unitaria real U se denomina *ortogonal*. Esta matriz se determina por las siguientes correlaciones:

$$UU' = U'U = E.$$

Todas las propiedades de las matrices ortogonales se deducen de las propiedades de las matrices unitarias.

Una matriz compleja cuadrada H se llama *hermitiana* o *autoconjugada*, si coincide con su inversa, es decir, si

$$H = H^*.$$

De este modo, la matriz de un operador hermitiano en la base ortonormalizada es hermitiana. De las propiedades del operador hermitiano se desprende que cualquier matriz hermitiana es semejante unitaria respecto a la matriz diagonal real. Si h_{ij} son los elementos de una matriz hermitiana H, entonces

$$h_{ij} = \overline{h}_{ji}$$

para cualesquiera i, j. De aquí obtenemos, en particular, que los elementos diagonales de cualquier matriz hermitiana son reales.

Una matriz hermitiana real H se denomina *simétrica*. Esta matriz se determina por la correlación siguiente:

$$H = H'$$

Observamos que toda matriz simétrica es semejante ortogonal respecto a la matriz diagonal real. Una matriz cuadrada se llama *normal*, si es conmutable con su inversa.

De acuerdo con esta definición, una matriz de un operador normal en la base ortonormalizada es normal. Teniendo presente las propiedades del operador normal, es fácil comprender que toda matriz normal compleja es semejante unitaria respecto de la matriz diagonal.

Las matrices de tipo especial son de mucha importancia en la construcción de los más diversos algoritmos de cálculo. No obstante, no serán objeto de estudio detallado. Todas las propiedades de estas matrices son, de hecho, la reflexión de las propiedades análogas de los operadores correspondientes.

Ejercicios.

1. Demuéstrese que toda matriz compleja es semejante unitaria respecto de la matriz triangular.

2. Sean $\lambda_1, \lambda_2, \ldots, \lambda_m$ los valores propios de la matriz A, con la particularidad de que cada valor propio se ha escrito tantas veces cual es su multiplicidad. Demuéstrese que

$$\sum_{i=1}^{m} |\lambda_i|^2 \leqslant \operatorname{tr} (A^*A). \qquad (81.1)$$

3. Demuéstrese que la igualdad en la correlación (81.1) tiene lugar cuando y sólo cuando, la matriz A es normal.

4. Haciendo uso de la fórmula de Binet—Cauchy, demuéstrese que para cualquier matriz A los menores principales de la matriz A^*A son no negativos.

5. Demuéstrese que la suma de cuadrados de los módulos de todos los menores de una matriz unitaria dispuestos en cualesquiera filas o columnas fijadas es igual a la unidad.

6. Demuéstrese que toda matriz rectangular A puede ser representada en la forma $A = Q\Lambda S$, donde Q, S son unas matrices unitarias y Λ es una matriz diagonal de elementos no negativos.

CAPÍTULO 10 PROPIEDADES MÉTRICAS DEL OPERADOR

§ 82. Continuidad y acotación del operador

Hemos introducido el concepto de operador lineal como cierta generalización del concepto de función. Si suponemos que en los espacios se ha definido cierta métrica, podemos observar la analogía con la acotación de una función, la continuidad de una función, etc. Al estudiar estos problemas, partiremos siempre de que un operador actúa del espacio normalizado m-dimensional X en el espacio normalizado n-dimensional Y. Si X no coincide con Y, las normas en ambos espacios pueden ser introducidas independientemente la una de la otra.

Un operador A que actúa de X en Y se denomina *continuo en el punto* $x_0 \in X$, si de la condición $x_k \to x_0$ se desprende que $Ax_k \to Ax_0$ para cualquier sucesión $\{x_k\}$ de X. Si el operador es continuo en todo punto del espacio X, se llama *continuo en todo punto* o, simplemente, *continuo*.

TEOREMA 82.1. *Un operador lineal que actúa en unos espacios normalizados arbitrarios de dimensión finita es continuo.*

DEMOSTRACIÓN. Tomemos un vector arbitrario $x_0 \in X$ y elijamos en X una base cualquiera $e_1, e_2, \ldots, e_m$. Tenemos

$$x_0 = \xi_1^{(0)} e_1 + \ldots + \xi_m^{(0)} e_m.$$

Supongamos que $x_k \to x_0$ y

$$x_k = \xi_1^{(k)} e_1 + \ldots + \xi_m^{(k)} e_m.$$

En conformidad con el teorema 53.1, de la convergencia en norma proviene la convergencia de coordenadas. Por ello, $\xi_s^{(k)} \to \xi_s^{(0)}$ para todo s. Pero

$$Ax_0 = \xi_1^{(0)} Ae_1 + \ldots + \xi_m^{(0)} Ae_m$$

y, además,

$$Ax_k = \xi_1^{(k)} Ae_1 + \ldots + \xi_m^{(k)} Ae_m.$$

Ahora, de la convergencia $\xi_s^{(k)} \to \xi_s^{(0)}$ para todos los s se deducirá la convergencia $Ax_k \to Ax_0$ en norma del espacio Y.

El operador A se llama *acotado*, si existe una constante M tal que $\| Ax \| \leqslant M \| x \|$ para todo vector $x \in X$.

TEOREMA 82.1. *Un operador lineal que actúa en unos espacios normalizados arbitrarios de dimensión finita es acotado.*

DEMOSTRACIÓN. Supongamos que en cierto caso el operador A no es acotado. Entonces, existe una sucesión de vectores no nulos $\{x_k\}$ tal que

$$\| Ax_k \| \geqslant k \| x_k \|.$$

Examinemos una sucesión de vectores

$$y_k = \frac{1}{k \| x_k \|} x_k.$$

Esta sucesión converge a cero, puesto que

$$\| y_k \| = \frac{1}{k \| x_k \|} \| x_k \| = \frac{1}{k} \to 0.$$

Por otra parte,

$$\| Ay_k \| = \frac{1}{k \| x_k \|} \| Ax_k \| \geqslant 1.$$

Esto es testimonio de que la sucesión $\{Ay_k\}$ no converge a cero, es decir, el operador A no es continuo en cero. La contradicción obtenida con el teorema 82.1 da por terminada la demostración.

Resulta natural plantear la cuestión acerca de la constante mínima de todas las constantes M que satisfacen la condición $\| Ax \| \leqslant M \| x \|$ para todos los vectores x. Puesto que el conjunto de estas constantes está acotado inferiormente por cero, la constante mínima existe a ciencia cierta. Se llama *norma del operador* A y se designa con el símbolo $\| A \|$. Por definición, la norma de un operador posee las siguientes dos propiedades:

1) para todo vector x del espacio X es válida la desigualdad

$$\| Ax \| \leqslant \| A \| \cdot \| x \|, \tag{82.1}$$

2) para todo número $\varepsilon > 0$ existe tal vector $x_\varepsilon \in X$ que

$$\| Ax_\varepsilon \| \geqslant (\| A \| - \varepsilon) \| x_\varepsilon \| \tag{82.2}$$

Demostremos que

$$\| A \| = \sup_{\| x \| \leqslant 1} \| Ax \| \tag{82.3}$$

o bien, que es lo mismo,

$$\| A \| = \sup_{x \neq 0} \frac{\| Ax \|}{\| x \|}; \tag{82.4}$$

si, desde luego, $\dim X > 0$.

Tomemos un vector arbitrario x que satisface la desigualdad $\| x \| \leqslant 1$. Entonces, de (82.1) se deduce que

$$\| Ax \| \leqslant \| A \| \, \| x \| \leqslant \| A \|.$$

Por consiguiente,

$$\sup_{\|x\|\leqslant 1} \|Ax\| \leqslant \|A\|. \qquad (82.5)$$

Luego, tomemos, de acuerdo con (82.2), un vector cualquiera x_ε y construyamos el vector

$$y_\varepsilon = \frac{1}{\|x_\varepsilon\|} x_\varepsilon.$$

En este caso

$$\|Ay_\varepsilon\| = \frac{1}{\|x_\varepsilon\|} \|Ax_\varepsilon\| \geqslant \frac{1}{\|x_\varepsilon\|} (\|A\| - \varepsilon) \|x_\varepsilon\| = \|A\| - \varepsilon.$$

Como $\|y_\varepsilon\| = 1$, resulta

$$\sup_{\|x\|\leqslant 1} \|Ax\| \geqslant \|Ay_\varepsilon\| \geqslant \|A\| - \varepsilon.$$

Por ser ε arbitrario, se obtiene que

$$\sup_{\|x\|\leqslant 1} \|Ax\| \geqslant \|A\|. \qquad (82.6)$$

Ahora, de (82.5), (82.6) se desprende la correlación (82.3) que se trataba de establecer.

Mostraremos a continuación que *la norma de un operador desempeña un papel excepcionalmente importante al introducir una métrica en los espacios de operadores lineales. En ese caso será esencial que la norma del operador tenga una forma explícita* (82.3).

Ejercicios

1. Demuéstrese que en un conjunto cerrado acotado de vectores se alcanzan las cotas superior e inferior de las normas de los valores del operador lineal.
2. Demuéstrese que un operador lineal transforma todo conjunto cerrado acotado en otro conjunto cerrado acotado.
3. ¿Será cierta la afirmación del ejercicio anterior, si no se requiere la acotación del conjunto?
4. Demuéstrese que en la fórmula (82.3) la cota superior se alcanza en un conjunto de vectores que satisfacen la condición $\|x\| = 1$, siempre que $\dim X > 0$.
5. Supongamos que un operador A actúa en el espacio X. Demuéstrese que A es regular, si, y sólo si, existe tal número $m > 0$ que $\|Ax\| \geqslant m\|x\|$ para todo $x \in X$.

§ 83. Norma del operador

Un conjunto ω_{XY} de operadores lineales que actúan de X en Y es un espacio lineal de dimensión finita. Si este espacio es lineal o complejo, se lo puede transformar en un espacio métrico completo, introduciendo en el primero, de tal o cual manera, una norma.

La introducción de la norma en un espacio de operadores lineales se efectúa mediante los mismos procedimientos que se usan en cual-

quier otro espacio lineal. No obstante, en el caso dado el mayor interés lo representan sólo aquellas normas en ω_{XY} que están relacionadas, de una manera suficientemente estrecha, con las normas en los espacios X, Y. Una de las clases más importantes de normas de este género lo constituyen las llamadas normas concordadas.

Si para todo operador de ω_{XY} se verifica la desigualdad

$$\| Ax \| \leqslant \| A \| \cdot \| x \|,$$

cualquiera que sea $x \in X$, la norma de los operadores se denomina *concordada* con las normas vectoriales en los espacios X, Y.

La ventaja de las normas concordadas se ve claramente en el siguiente ejemplo. Supongamos que λ es el valor propio del operador A que actúa en el espacio X, y x, un vector propio correspondiente a λ. En este caso $Ax = \lambda x$, y, por lo tanto,

$$| \lambda | \cdot \| x \| = \| \lambda x \| = \| Ax \| \leqslant \| A \| \cdot \| x \|.$$

Por consiguiente, $| \lambda | \leqslant \| A \|$. Así pues, hemos obtenido una deducción muy importante:

Los módulos de los valores propios de un operador lineal no son superiores a cualquiera de sus normas concordadas.

El ejemplo citado muestra que para obtener las mejores estimaciones, es deseable emplear la menor de las normas concordadas. Está claro que todas las normas concordadas están acotadas inferiormente por la expresión (82.3). Si probamos que esta expresión satisface los axiomas de la norma, ella será precisamente la menor de las normas concordadas. De este modo *justificaremos* tanto la denominación de la expresión (82.3), como su designación.

Evidentemente, para cualquier operador A la expresión $\| A \|$ es no negativa. Si $\| A \| = 0$, es decir, si

$$\sup_{\|x\| \leqslant 1} \| Ax \| = 0,$$

entonces $\| Ax \| = 0$ para todos los vectores x cuya norma no es superior a la unidad. Pero, en este caso, en virtud de la linealidad del operador, $Ax = 0$ para cualquier x. Por consiguiente $A = 0$. Para todo operador A y un número λ se tiene

$$\| \lambda A \| = \sup_{\|x\| \leqslant 1} \| \lambda Ax \| = | \lambda | \sup_{\|x\| \leqslant 1} \| Ax \| = | \lambda | \cdot \| A \|.$$

Y, por fin, para cualesquiera dos operadores A, B, de ω_{XY}

$$\| A + B \| = \sup_{\|x\| \leqslant 1} \| Ax + Bx \| \leqslant \sup_{\|x\| \leqslant 1} (\| Ax \| + \| Bx \|) \leqslant$$

$$\leqslant \sup_{\|x\| \leqslant 1} \| Ax \| + \sup_{\|x\| \leqslant 1} \| Bx \| = \| A \| + \| B \|.$$

Todas estas correlaciones significan que la expresión (82.3) representa en sí una norma en el espacio de operadores lineales. La norma (82.3)

se denomina norma del operador *subordinada* a las normas vectoriales en los espacios X, Y.

La norma subordinada posee también una propiedad muy importante con relación a la operación de multiplicación de los operadores. Supongamos que el operador A actúa de X en Y y el operador B, de Y en Z. Entonces, como se sabe, queda definido el operador BA. Teniendo presente la concordancia de las normas subordinadas, encontramos

$$\|BA\| = \sup_{\|x\|\leqslant 1} \|(BA)x\| = \sup_{\|x\|\leqslant 1} \|B(Ax)\| \leqslant$$
$$\leqslant \sup_{\|x\|\leqslant 1} (\|B\|\cdot\|Ax\|) = \|B\| \sup_{\|x\|\leqslant 1} \|Ax\| = \|B\|\cdot\|A\|.$$

De este modo, cualquier norma subordinada del operador posee las siguientes cuatro propiedades principales. Para cualesquiera operadores A, B y todo número λ

$$\begin{aligned} &1)\ \|A\| > 0, \text{ si } A \neq 0;\ \|0\| = 0,\\ &2)\ \|\lambda A\| = |\lambda|\,\|A\|,\\ &3)\ \|A + B\| \leqslant \|A\| + \|B\|,\\ &4)\ \|BA\| \leqslant \|B\|\cdot\|A\|. \end{aligned} \qquad (83.1)$$

Como propiedad adicional indiquemos que para un operador idéntico E es válida la igualdad

5) $\|E\| = 1$.

Esta se desprende de (82.3), puesto que $Ex = x$ para todo vector x.

En el caso general, la norma subordinada de un operador depende tanto de la norma en el espacio X, como también de la norma en el espacio Y. Si los dos espacios son unitarios, a título de norma en ellos puede servir la longitud de los vectores. La correspondiente norma subordinada del operador se denomina norma *espectral* y se designa con el símbolo $\|\cdot\|_2$. Así pues, para todo operador A que actúa de X en Y se tiene

$$\|A\|_2^2 = \sup_{(x,\,x)\leqslant 1} (Ax,\ Ax). \qquad (83.2)$$

Investiguemos algunas propiedades de la norma espectral.

La norma espectral no varía, cuando el operador se multiplica por cualesquiera operadores unitarios.

Sean V, U los operadores unitarios arbitrarios que actúan en los espacios X, Y, respectivamente. Consideremos el operador $B = UAV$. Tenemos

$$\|B\|_2^2 = \sup_{(x,\,x)\leqslant 1} (Bx,\ Bx) = \sup_{(x,\,x)\leqslant 1} (UAVx,\ UAVx) =$$
$$= \sup_{(x,\,x)\leqslant 1} (AVx,\ U^*UAVx) = \sup_{(x,\,x)\leqslant 1} (AVx,\ AVx) =$$
$$= \sup_{(Vx,\,Vx)\leqslant 1} (AVx,\ AVx) = \sup_{(v,\,v)\leqslant 1} (Av,\ Av) = \|A\|_2^2,$$

La definición de una norma espectral en la forma de (83.2) permite establecer su relación con los números singulares del operador A. Sean $x_1, x_2, \ldots, x_m$ un sistema ortonormalizado de los vectores propios del operador A^*A y sean $\rho_1^2, \rho_2^2, \ldots, \rho_m^2$ los valores propios de dicho operador. Sin limitar la generalidad de los razonamientos supongamos que

$$\rho_1 \geqslant \rho_2 \geqslant \ldots \geqslant \rho_m \geqslant 0. \qquad (83.3)$$

Representemos el vector $x \in X$ en forma de la descomposición

$$x = \alpha_1 x_1 + \ldots + \alpha_m x_m, \qquad (83.4)$$

entonces

$$(x, x) = \sum_{i=1}^{m} |\alpha_i|^2.$$

Según se ha observado en el § 78, el sistema $x_1, x_2, \ldots, x_m$ se transforma por el operador A en un sistema ortogonal, siendo en este caso

$$(Ax_i, Ax_i) = \rho_i^2$$

para todo i. Por consiguiente,

$$(Ax, Ax) = \sum_{i=1}^{m} |\alpha_i|^2 \rho_i^2,$$

que da

$$\|A\|_2^2 = \sup_{\sum_{i=1}^{m} |\alpha_i|^2 \leqslant 1} \sum_{i=1}^{m} |\alpha_i|^2 \rho_i^2. \qquad (83.5)$$

Está claro que bajo las condiciones (83.3)

$$\|A\|_2^2 \leqslant \rho_1^2.$$

Pero, para el vector x_1 el segundo miembro de (83.5) toma el valor ρ_1^2. Por ello

$$\|A\|_2^2 = \rho_1^2.$$

De este modo,

La norma espectral del operador A es igual al número singular máximo.

Recordemos que para un operador normal A los números singulares coinciden con los módulos de los valores propios. Por consiguiente, la norma espectral del operador unitario es igual a la unidad, la norma espectral de un operador no negativo es igual al valor propio máximo.

Ejercicios.

1. Demuéstrese que para cualquier valor propio λ del operador A se verifica la desigualdad

$$|\lambda| \leqslant \inf_k \|A^k\|^{1/k}.$$

2. Sea $\varphi(z)$ cualquier polinomio con coeficientes no negativos. Demuéstrese que

$$\| \varphi(A) \| \leqslant \varphi(\| A \|).$$

3. Demuéstrese que $\| A \| \geqslant \| A^{-1} \|^{-1}$ para todo operador regular A. ¿Cuándo tendrá lugar la igualdad para el caso de una norma espectral?

§ 84. Normas matriciales del operador

La norma espectral es, en esencia, la única norma subordinada del operador el cálculo de la cual no está relacionado explícitamente con las bases. En cambio, si en los espacios con operadores dados se han fijado algunas bases, entonces la posibilidad para introducir las normas operacionales se hace mucho más amplia.

Así pues, consideraremos una vez más los operadores lineales que actúan del espacio X en el espacio Y. Supongamos que en X está fijada la base $e_1, e_2, \ldots, e_m$ y en Y, la base $q_1, q_2, \ldots, q_n$. Descompongamos un vector arbitrario $x \in X$ según la base y obtendremos

$$x = x_1 e_1 + \ldots + x_m e_m. \tag{84.1}$$

Ahora, la norma en el espacio X puede introducirse, por ejemplo, de acuerdo con la fórmula (52.3) o por cualquier otro método, sirviéndose de los coeficientes de la descomposición. De modo análogo puede ser introducida también una norma en el espacio Y.

Las más usadas son las normas del tipo (52.4). Por ello, investigaremos las normas de los operadores subordinadas y concordadas precisamente con las del tipo (52.4). Más aún, convengamos en considerar que en ambos espacios X e Y se han introducido las normas de un mismo tipo. Es evidente que las normas correspondientes del operador A han de ser ligadas de tal o cual manera con los elementos a_{ij} de la matriz del operador en las bases elegidas.

Demos a conocer, al principio, las expresiones para las normas de los operadores subordinadas a las 1-normas y ∞-normas de (52.4). Tenemos

$$\| A \|_1 = \sup_{\| x \|_1 \leqslant 1} \| Ax \|_1 = \sup_{\| x \|_1 \leqslant 1} \Big(\sum_{i=1}^{n} \Big| \sum_{j=1}^{m} a_{ij} x_j \Big| \Big) \leqslant$$

$$\leqslant \sup_{\| x \|_1 \leqslant 1} \Big(\sum_{i=1}^{n} \sum_{j=1}^{m} | a_{ij} | \, | x_j | \Big) = \sup_{\| x \|_1 \leqslant 1} \Big(\sum_{j=1}^{m} | x_j | \sum_{i=1}^{n} | a_{ij} | \Big) \leqslant$$

$$\leqslant \Big(\max_{1 \leqslant j \leqslant m} \sum_{i=1}^{n} | a_{ij} | \Big) \Big(\sup_{\| x \|_1 \leqslant 1} \| x \|_1 \Big) = \max_{1 \leqslant j \leqslant m} \sum_{i=1}^{n} | a_{ij} |.$$

Mostremos ahora que para cierto vector x que satisface la condición $\| x \|_1 \leqslant 1$, $\| Ax \|_1$ coincide con el segundo miembro de la correlación obtenida.

Supongamos que el valor máximo en el segundo miembro se alcanza cuando $j = l$. En este caso todas las desigualdades se convierten en igualdades, por ejemplo, para $x = e_l$. Así pues,

$$\|A\|_1 = \max_{1\leqslant j\leqslant n} |a_{ij}|.$$

Análogamente se investiga también la otra norma:

$$\|A\|_\infty = \sup_{\|x\|_\infty\leqslant 1} \|Ax\|_\infty = \sup_{\|x\|_\infty\leqslant 1} \Big(\max_{1\leqslant i\leqslant m} \Big|\sum_{j=1}^{n} a_{ij}x_j\Big|\Big) \leqslant$$

$$\leqslant \sup_{\|x\|_\infty\leqslant 1} \Big(\max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|\,|x_j|\Big) \leqslant$$

$$\leqslant \sup_{\|x\|_\infty\leqslant 1} \Big(\Big(\max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|\Big)\Big(\max_{1\leqslant j\leqslant n} |x_j|\Big)\Big) =$$

$$= \Big(\max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|\Big)\Big(\sup_{\|x\|_\infty\leqslant 1} \|x\|_\infty\Big) = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|.$$

Supongamos que el valor máximo en el segundo miembro se alcanza para $i = l$. Tomemos un vector x cuyas coordenadas son $x_j = |a_{lj}|/a_{lj}$, si $a_{lj} \neq 0$ y $x_j = 1$, si $a_{lj} = 0$. No es difícil comprobar que para dicho vector todas las desigualdades se convierten en igualdades. Por consiguiente,

$$\|A\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|.$$

Con el objeto de hallar la norma del operador subordinada a las 2-normas de (52.4) procedamos de la manera siguiente. Por analogía con (32.1), introduzcamos un producto escalar en los espacios X, Y. Entonces la 2-norma de (52.4) coincidirá con la longitud del vector. Por ello, la norma subordinada no es otra cosa que la norma espectral del operador, correspondiente al producto escalar dado. Elegidos los productos escalares, las bases se convierten en ortonormalizadas, razón por la cual en estas bases al operador conjugado le corresponderá una matriz conjugada. Al designar mediante A_{qe} la matriz del operador A, obtendremos de lo dicho la siguiente deducción.

La norma de un operador subordinada a las 2-normas es igual al número singular máximo de la matriz A_{qe}.

Las normas examinadas son ciertas funciones de la matriz del operador. Mediante un procedimiento semejante pueden construirse no sólo las normas subordinadas, sino también las concordadas. Una de las más importantes normas concordadas es la llamada norma *euclidiana*. Designarémosla con el símbolo $\|\cdot\|_E$. Si el operador A tiene en las bases elegidas la matriz A_{qe} con elementos a_{ij}, entonces,

por definición,

$$\|A\|_E = \Big(\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2\Big)^{1/2}.$$

El segundo miembro de esta expresión es una norma en el espacio $n \times m$-dimensional de operadores lineales. Por ello, el cumplimiento de las primeras tres propiedades de (83.1) no causa duda alguna. Es de mucha importancia el hecho de que para la norma euclidiana se cumple también la cuarta propiedad de (83.1). Para demostrar esto haremos uso de la desigualdad de Cauchy — Buniakovski del tipo (27.5).

Consideraremos los espacios lineales X, Y, Z de dimensiones respectivas m, n, p. Supongamos que el operador A actúa de X en Y, y el operador B, de Y en Z. Designamos mediante a_{ij}, b_{ij} los elementos de las matrices de estos operadores en las bases elegidas. Tenemos

$$\|BA\|_E = \Big(\sum_{i=1}^{p} \sum_{j=1}^{m} \Big|\sum_{k=1}^{n} b_{ik}a_{kj}\Big|^2\Big)^{1/2} \leqslant \Big(\sum_{i=1}^{p} \sum_{j=1}^{m} \Big(\sum_{k=1}^{n} |b_{ik}|\,|a_{kj}|\Big)^2\Big)^{1/2} \leqslant$$

$$\leqslant \Big(\sum_{i=1}^{p} \sum_{j=1}^{m} \Big(\sum_{k=1}^{n} |b_{ik}|^2\Big)\Big(\sum_{k=1}^{n} |a_{kj}|^2\Big)\Big)^{1/2} =$$

$$= \Big(\Big(\sum_{i=1}^{p} \sum_{k=1}^{n} |b_{ik}|^2\Big)\Big(\sum_{k=1}^{n} \sum_{j=1}^{m} |a_{kj}|^2\Big)\Big)^{1/2} = \|B\|_E \cdot \|A\|_E.$$

En el caso general la norma euclidiana no es subordinada. El hecho de que ella esté compatible con las 2-normas se demuestra del mismo modo que el utilizado para demostrar la propiedad que acabamos de considerar.

La comprobación directa permite establecer una fórmula de importancia para la norma euclidiana. A saber,

$$\|A\|_E^2 = \operatorname{tr}(A_{qe}^{*}A_{qe}) = \operatorname{tr}(A_{qe}A_{qe}^{*}). \qquad (84.2)$$

Ahora podemos enunciar las siguientes deducciones.

A una matriz conjugada en las bases ortonormalizadas le corresponde un operador conjugado. Al introducir en los espacios X, Y los productos escalares por analogía con (32.1), convertiremos las bases elegidas en las ortonormalizadas. Dado que la traza de una matriz es igual a la suma de sus valores propios de (84.2) se deduce que:

El cuadrado de la norma euclidiana de un operador es igual a la suma de los cuadrados de sus números singulares.

Al introducir productos escalares en X, Y podemos hablar de los operadores unitarios. Con referencia a estos operadores unitarios resulta fácil mostrar que:

La norma euclidiana no varía al multiplicar el operador por cualesquiera operadores unitarios.

En efecto, como ya se ha indicado en los ejercicios del § 78, los números singulares no cambian cuando se multiplican por los operadores unitarios, mientras que la norma euclídiana se expresa solamente en términos de los números singulares.

En la mayoría de las aplicaciones relacionadas con las normas resulta importante no tanto la definición explícita de la norma del operador, como el cumplimiento de las propiedades (83.1). Por esta razón la norma de un operador puede definirse *axiomáticamente* por intermedio de su matriz. Elijamos en los espacios, donde están fijados unos operadores, algunas bases; entonces a todo operador corresponderá cierta matriz. A toda matriz pondremos en correspondencia un número denotado por el símbolo $\|\cdot\|$ y supongamos que en este caso quedan cumplidas como axiomas las condiciones (83.1). El número $\|\cdot\|$ se llamará *norma de la matriz*. Si ahora a todo operador se le pone en correspondencia la norma de su matriz, será obvio que de este modo en el espacio de operadores se introduce una norma. Es evidente que las condiciones (83.1) se cumplen también para los operadores. Lo recíproco es también cierto. Toda norma del operador engendra, con bases fijadas, una norma de la matriz. Estas normas de las matrices se designarán mediante símbolos análogos $\|\cdot\|_2$, $\|\cdot\|_\infty$, etc. Por lo visto, la concordancia de la norma también puede exigirse axiomáticamente.

Los ejemplos examinados muestran que la realización práctica de la definición axiomática de una norma del operador en términos de la norma de la matriz es posible. En lo que sigue, al hablar de las normas de las matrices y de los operadores, *siempre* supondremos su concordancia y el cumplimiento de las condiciones (83.1).

Ejercicios.

1. Demuéstrese que con toda norma para la matriz unidad se verifica la desigualdad

$$\|E\| \geq 1. \tag{84.3}$$

2. Sean $\lambda_1, \ldots, \lambda_n$ los valores propios de la matriz A. Demuéstrese que

$$\inf_B \|B^{-1}AB\|_E^2 = \sum_{k=1}^{n} |\lambda_k|^2.$$

Compárese esta igualdad con (81.4).

§ 85. Ecuaciones operacionales

Uno de los más importantes problemas del álgebra consiste en la resolución de los sistemas de ecuaciones algebraicas lineales. Ya nos hemos encontrado más de una vez con este problema en el transcurso de nuestra narración. Ahora lo

consideraremos desde el punto de vista de la teoría de operadores lineales.

Supongamos que está dado un sistema (60.2) con elementos del campo P de números reales o complejos. Tomemos un espacio m-dimensional X y un espacio n-dimensional Y sobre un mismo campo P y fijamos en ellos unas bases cualesquiera. En este caso las correlaciones (60.2) serán equivalentes a una igualdad matricial del tipo (61.2) y la última es, a su vez, equivalente a la igualdad operacional

$$Ax = y. \tag{65.1}$$

Aquí, A es un operador que actúa de X en Y y tiene en las bases elegidas la misma matriz de que se dispone el sistema (60.2). Las coordenadas de los vectores $x \in X$ e $y \in Y$ en las bases elegidas son, respectivamente, $(\xi_1, \ldots, \xi_m)$ y $(\eta_1, \ldots, \eta_n)$.

De este modo, en lugar de los sistemas de ecuaciones algebraicas lineales podemos considerar las ecuaciones (65.1). El problema consiste en hallar todos los vectores $x \in X$ que para el operador A y el vector $y \in Y$ dados satisfacen la igualdad (65.1). Una ecuación del tipo (65.1) se denomina *operacional*, el vector y lleva el nombre de *segundo miembro* y el vector x es *la solución*. Desde luego, todas las propiedades de los sistemas de ecuaciones se extienden automáticamente a las ecuaciones operacionales, y viceversa.

El teorema de Kronecker—Capelli enuncia la condición necesaria y suficiente para que el sistema sea resoluble en términos del rango de la matriz. Esto no es muy cómodo, porque no permite observar una relación profunda que existe entre los sistemas y las ecuaciones de otros tipos.

Sean X, Y unos espacios unitarios, entonces queda definido el operador A^*. La ecuación (65.1) se llamará *ecuación no homogénea fundamental* y la ecuación

$$A^*u = 0,$$

ecuación no homogénea conjugada. Si los segundos miembros son nulos, las ecuaciones correspondientes se denominarán *homogéneas*. Resulta lícita la siguiente afirmación:

O bien la ecuación no homogénea fundamental tiene solución, cualquiera que sea el segundo miembro, o bien la ecuación homogénea conjugada tiene por lo menos una solución no nula.

En efecto, designemos con r al rango del operador A. El mismo rango tendrá también el operador A^*. Pueden observarse dos casos: o bien $r = n$, o bien $r < n$. En el primer caso el campo de valores del operador A tiene dimensión n y, por lo tanto, coincide con el espacio Y. Por eso la ecuación no homogénea fundamental debe tener solución, cualquiera que sea el segundo miembro. En el mismo caso el defecto del operador conjugado es igual a cero, razón por la cual el núcleo no tiene vectores no nulos, es decir, la ecuación homogénea

conjugado no tiene soluciones no nulas. Si $r < n$, el campo de valores del operador A no coincide con Y y la ecuación no homogénea fundamental no puede tener solución, cualquiera que sea el segundo miembro. En esta circunstancia el núcleo del operador conjugado se compone no sólo del vector nulo, a consecuencia de lo cual la ecuación conjugada homogénea tiene soluciones no nulas.

La afirmación demostrada es de significación especial cuando los espacios X, Y coinciden. En este caso la existencia de una solución de la ecuación no homogénea fundamental, cualquiera que sea el segundo miembro, significa que el operador A es regular. Por ello, en el caso dado es válida la así llamada

ALTERNATIVA DE FREDHOLM. *O bien la ecuación no homogénea fundamental tiene siempre solución y, además, única, cualquiera que sea el segundo miembro o bien la ecuación homogénea conjugada tiene por lo menos una solución no nula.*

TEOREMA DE FREDHOLM. *Para que la ecuación no homogénea fundamental sea resoluble, es necesario y suficiente que su segundo miembro sea ortogonal a todas las soluciones de la ecuación homogénea conjugada.*

DEMOSTRACIÓN. Designemos con N^* el núcleo del operador A^* y mediante T, el campo de valores del operador A. Si la ecuación no homogénea fundamental es resoluble, entonces el segundo miembro $y \in T$. De acuerdo con (75.8), se infiere que $y \perp N^*$, es decir, $(y, u) = 0$ para todos los vectores u que satisfacen la ecuación $A^*u = 0$. Sea ahora $(y, u) = 0$ para los mismos vectores u, entonces $y \perp N^*$ y, conforme a (75.8), $y \in T$. Mas, esto significa que existe un vector $x \in X$ tal que $Ax = y$, es decir, la ecuación no homogénea fundamental es resoluble.

Ejercicios.

1. Demuéstrese que la ecuación $A^*Ax = A^*y$ es resoluble.
2. Demuéstrese que la ecuación $(A^*A)^p x = (A^*A)^q y$ es resoluble para cualesquiera p, q enteros y positivos.
3. Establézcase el sentido geométrico de la alternativa y del teorema de Fredholm.

§ 86. Seudosoluciones y un operador seudoinverso

La definición arbitraria del operador A y del miembro segundo y puede conducir a que la ecuación (85.1) no tenga ninguna solución. Evidentemente, esto se debe sólo a *qué se entiende exactamente por solución de una ecuación.*

Tomemos un vector arbitrario $x \in X$ y consideremos el vector $r = Ax - y$, llamado *residuo* del vector x. Para que x sea la solución de la ecuación (85.1), es necesario y suficiente que su residuo sea

nulo. A su vez, para que un residuo sea nulo, es necesario y suficiente que sea nula su longitud. De este modo, todas las soluciones de la ecuación (85.1), y sólo ellas, satisfacen la igualdad

$$|Ax - y|^2 = 0.$$

Dado que el valor nulo de la longitud del residuo es el mínimo, la determinación de las soluciones de la ecuación (85.1) puede considerarse como el problema de la búsqueda de los vectores x, para los cuales la expresión

$$\Phi_y(x) = |Ax - y|^2 \qquad (86.1)$$

alcanza su valor mínimo. El segundo miembro de esta expresión se denomina *funcional del residuo*. La búsqueda de los vectores que minimizan la funcional del residuo tiene también sentido en el caso cuando las soluciones de la ecuación (85.1) no existan. Esto sirve de base para la siguiente definición.

Se llama *seudosolución* (o *solución generalizada*) de la ecuación (85.1) todo vector $x \in X$, para el cual la funcional del residuo alcanza su valor mínimo. La seudosolución de longitud mínima recibe el nombre de seudosolución *normal*.

Mostremos que la seudosolución normal existe siempre y es, además, única. En los espacios X, Y fijemos las bases singulares $x_1, \ldots, x_m$ e $y_1, \ldots, y_n$. Sea

$$x = \sum_{k=1}^{m} \alpha_k x_k, \qquad y = \sum_{p=1}^{n} \beta_p y_p. \qquad (86.2)$$

Tomando en consideración las correlaciones (78.2), obtenemos

$$Ax - y = \sum_{k=1}^{m} \rho_k \alpha_k y_k - \sum_{p=1}^{n} \beta_p y_p.$$

Convengamos en considerar, como hasta ahora, que los números singulares $\rho_1, \ldots, \rho_r$ son distintos de cero, mientras que los números restantes son nulos. Puesto que las bases singulares son ortonormalizadas, tenemos

$$\Phi_y(x) = \sum_{k=1}^{r} |\rho_k \alpha_k - \beta_k|^2 + \sum_{p=r+1}^{n} |\beta_p|^2.$$

Es evidente que el valor mínimo de la funcional del residuo se conseguirá para aquellos vectores x, en los cuales las últimas $m - r$ coordenadas α_k son arbitrarias, y las primeras r coordenadas se determinan por la fórmula

$$\alpha_k = \beta_k/\rho_k. \qquad (86.3)$$

La seudosolución normal será

$$x_0 = \sum_{k=1}^{r} \frac{\beta_k}{\rho_k} x_k. \qquad (86.4)$$

Recordemos que los vectores $x_{t+1}, \ldots, x_m$ forman la base del núcleo N del operador A. Por eso el conjunto de todas las seudosoluciones representa en sí un plano en el espacio X cuyo subespacio director coincide con el núcleo N y el vector de desplazamiento coincide con cualquier seudosolución. *La solución seudonormal es el único vector, ortogonal a N, de este plano.*

Haciendo uso de las correlaciones (78.2), (78.3), es fácil mostrar que las seudosoluciones, y sólo ellas, satisfacen la ecuación

$$A^*Ax = A^*y. \tag{86.5}$$

En efecto, escribamos los vectores x, y en forma de las descomposiciones (86.2). Tenemos

$$A^*Ax = \sum_{k=1}^{t} \rho_k^2 \alpha_k x_k, \quad A^*y = \sum_{p=1}^{t} \rho_p \beta_p x_p.$$

De aquí se deduce que las soluciones de la ecuación (86.5) serán los vectores x, y sólo ellos, para los cuales las primeras t coordenadas α_k se calculan según (86.3) y las últimas $m - t$ coordenadas son arbitrarias.

De este modo, *si la resolubilidad de la ecuación* (85.1) *no es garantizada, siempre podemos sustituir la resolución de dicha ecuación por la resolución de la ecuación* (86.5). *En este caso se asegura la minimización de la funcional del residuo para la ecuación* (85.1).

El operador inverso juega un papel importante en muchas investigaciones. No obstante, fue definido sólo para un operador regular y por ahora no disponemos del análogo correspondiente para un operador degenerado y un operador que actúa de un espacio en otro. Dicho análogo puede ser construido sobre la base de seudosoluciones.

Supongamos que el operador A actúa del espacio X en el espacio Y. En este caso, a todo vector $y \in Y$ podemos ponerle en correspondencia un único vector $x_0 \in X$ que representa la seudosolución normal de la ecuación (85.1). Dicha correspondencia determina cierto operador A^+, el cual actúa de Y en X y lleva el nombre de operador *seudoinverso* (o *inverso generalizado*) del operador A. Por consiguiente, según la definición,

$$x_0 = A^+y \tag{86.6}$$

para cualquier $y \in Y$. Está claro que si el operador A es regular, el operador *seudoinverso coincide con el inverso*. Investiguemos las propiedades del operador seudoinverso.

Supongamos que a la par con (86.6) disponemos de $u_0 = A^+v$ para cierto vector $v \in Y$. Consideremos un vector $\alpha y + \beta v$ para cualesquiera números α, β. Al tomarlo en calidad de segundo miembro de la ecuación (85.1), el vector $\alpha x_0 + \beta u_0$ satisfará, a ciencia cierta, una ecuación correspondiente del tipo (86.5) y por esta razón será una seudosolución. Puesto que x_0, u_0 son ortogonales al núcleo del

operador A, será también ortogonal al núcleo del vector $\alpha x_0 + \beta u_0$. Por lo tanto, dicho vector es la seudosolución normal. De este modo, *la linealidad* del operador seudoinverso queda establecida.

Las propiedades del operador seudoinverso pueden ser fácilmente establecidas, si consideramos su acción sobre los vectores de las bases singulares. De acuerdo con (86.4), tenemos

$$A^+ y_k = \begin{cases} \rho_k^{-1} x_k, & k \leqslant r, \\ 0, & k > r. \end{cases} \qquad (86.7)$$

De aquí se desprende que:

El campo de definición, el núcleo y el campo de valores de los operadores seudoinverso y conjugado coinciden.

Con ayuda de las fórmulas (78.2), (78.3), (86.7) se pueden obtener diferentes correlaciones que relacionan los operadores A, A^*, A^+. He aquí algunas de ellas:

1) $(A^*)^+ = (A^+)^*$,

2) $(A^+)^+ = A$,

3) $(AA^+)^* = AA^+$, $\quad (AA^+)^2 = AA^+$,

4) $(A^+A)^* = A^+A$, $\quad (A^+A)^2 = A^+A$,

5) $AA^+A = A$.

Estas correlaciones se demuestran según el mismo esquema, razón por la cual a título de ejemplo consideraremos detalladamente sólo las correlaciones primera y tercera.

Comparando (78.2) y (86.7) elijamos, en calidad del operador A, el operador conjugado A^*. Puesto que para este último se verifica (78.3), resulta

$$(A^*)^+ x_k = \begin{cases} \rho_k^{-1} y_k, & k \leqslant r, \\ 0, & k > r. \end{cases}$$

Ahora, partiendo de (86.7), apliquemos una correlación, análoga a (78.3), al operador $(A^+)^*$. En este caso

$$(A^+)^* x_k = \begin{cases} \rho_k^{-1} y_k, & k \leqslant r, \\ 0, & k > r. \end{cases}$$

De este modo, los operadores $(A^*)^+$ y $(A^+)^*$ coinciden sobre la base $x_1, \ldots, x_n$, y, por tanto, son iguales.

Al tomar en consideración (78.2), (86.7), concluimos que para el operador AA^+ son válidas las correlaciones

$$AA^+ y_k = \begin{cases} 1 \cdot y_k, & k \leqslant r, \\ 0, & k > r. \end{cases} \qquad (86.8)$$

Esto significa que el operador AA^+ tiene un sistema ortonormalizado de vectores propios $y_1, \ldots, y_m$, así como también los valores propios

reales 1 y 0, es decir, es hermitiano. De este modo queda establecida la primera de las correlaciones del tercer grupo. La segunda igualdad se desprende, evidentemente, de (86.8).

Ejercicios.

1. ¿Qué representa en sí un operador que es seudoinverso respecto al operador nulo?

2. Supongamos que los espacios X, Y son distintos. Escríbase la matriz de un operador seudoinverso en las bases singulares y compárese ésta con (78.4).

3. Sean U, V unos operadores unitarios que actúan en los espacios X e Y, respectivamente. Demuéstrese que

$$(VAU)^+ = U^*A^+V^*.$$

4. Demuéstrese que existen tales operadores K en X y L en Y que se verifica

$$A^+ = KA^* = A^*L.$$

Descríbase la acción de los operadores K, L.

5. Demuéstrese que un operador seudoinverso se define unívocamente por las condiciones

$$AA^+A = A,$$

$$A^+ = KA^* = A^*L.$$

6. Demuéstrese que todas las seudosoluciones, y sólo ellas, sirven de soluciones para la ecuación

$$Ax = AA^+y.$$

7. Establézcase el sentido geométrico de las seudosoluciones.

§ 87. Perturbación y regularidad del operador

Hemos subrayado más de una vez que una variación pequeña de una base, las coordenadas de un vector o de los elementos de una matriz, etc. puede tener por resultado el cambio de varias propiedades relacionadas con la noción de dependencia lineal. Esta noción desempeña un papel decisivo en toda la teoría de operadores lineales, a consecuencia de lo cual resulta muy importante investigar *la influencia de la pequeña variación* de los operadores en las propiedades de los mismos.

En la resolución de los más diversos problemas hemos de utilizar, como un medio auxiliar, un operador *próximo al operador idéntico*. Por este término se entenderá un operador que actúa en el espacio X y tiene la forma $E + A$, donde $\|A\| < 1$ para una de las normas.

Si λ es un valor propio cualquiera del operador A, entonces $1 + \lambda$ será un valor propio del operador $E + A$. Dado que $|\lambda| \leqslant \|A\|$, entonces, en virtud de la condición $\|A\| < 1$, todos los valores propios del operador A son inferiores en módulo a la unidad. Por consiguiente, todos los valores propios del operador $E + A$ son distintos de cero y este operador será regular.

De este modo, *al cumplirse la condición* $\|A\| < 1$, *existe el operador* $(E + A)^{-1}$. *En cambio, si el operador* $E + A$ *es degenerado, entonces* $\|A\| \geqslant 1$ *para cualquier norma.*

Para cualquier número α, cuyo módulo es inferior a la unidad, es válida la correlación límite

$$(1+\alpha)^{-1} = \lim_{p\to\infty} \alpha_p$$

donde

$$\alpha_p = \sum_{k=0}^{p} (-\alpha)^k.$$

Mostremos que una correlación análoga tiene lugar también para el operador $(E + A)^{-1}$, si $\|A\| < 1$. Examinemos una sucesión de operadores $\{A_p\}$

$$A_p = \sum_{k=0}^{p} (-A)^k.$$

Es fácil comprobar que

$$(E + A) A_p = E - (-A)^{p+1},$$

por lo cual

$$\|(E + A) A_p - E\| = \|A^{p+1}\|. \tag{87.1}$$

Formalmente esta igualdad es verídica también para $p = -1$, siempre que se considere $A_{-1} = 0$.

Luego, se tiene

$$\|(E + A) A_p - E\| = \|(A_p - (E + A)^{-1} + A(A_p-(E+A)^{-1})\| \geqslant$$
$$\geqslant |\,\|A_p - (E + A)^{-1}\| - \|A\|\cdot\|A_p - (E+A)^{-1}\|\,| =$$
$$= (1 - \|A\|)\,\|A_p - (E + A)^{-1}\|.$$

Ahora, teniendo presente (87.1), obtenemos, para $p = -1$, la estimación de la norma del operador $(E + A)^{-1}$, es decir

$$\|(E + A)^{-1}\| \leqslant \frac{\|E\|}{1-\|A\|}.$$

Para cualquier norma subordinada se verifica la igualdad $\|E\| = 1$, por consiguiente, en este caso

$$\|(E + A)^{-1}\| \leqslant \frac{1}{1-\|A\|}. \tag{87.2}$$

Cuando $p \geqslant 0$, obtenemos la estimación para la desviación del operador A_p del operador $(E + A)^{-1}$. A saber,

$$\|A_p - (E + A)^{-1}\| \leqslant \frac{\|A\|^{p+1}}{1-\|A\|}. \tag{87.3}$$

En virtud de la condición $\|A\| < 1$, esto significa que la sucesión $\{A_p\}$ convergerá al operador $(E + A)^{-1}$. Si el operador A_p se consi-

dera como *aproximación* hacia el operador $(E + A)^{-1}$, la fórmula (87.3) nos ofrece la *estimación de la exactitud con que se realiza la aproximación*.

Sea A un operador regular cualquiera. Consideraremos el operador $A + \varepsilon_A$, donde ε_A es un operador arbitrario. Llamaremos ε_A *perturbación del operador* A, y $A + \varepsilon_A$, operador perturbado. Aclaremos en qué condiciones, impuestas sobre la magnitud de la norma de perturbación, el operador perturbado será regular. Nos serán de interés en este caso sólo los valores *pequeños* de la norma de perturbación.

El operador A es regular y por eso existe el operador A^{-1}. Por consiguiente, se verifica la igualdad

$$A + \varepsilon_A = A\,(E + A^{-1}\varepsilon_A).$$

De aquí se desprende que el operador $A + \varepsilon_A$ será regular, si, y sólo si, es regular el operador $E + A^{-1}\varepsilon_A$. Esta condición se cumple a ciencia cierta, siempre que

$$\| A^{-1}\varepsilon_A \| < 1$$

para una norma cualquiera. Se cumple con mayor razón, si $\| A^{-1} \| \, \| \varepsilon_A \| < 1$.

De suerte que *un operador perturbado será regular con todas perturbaciones que satisfacen la desigualdad*

$$\| \varepsilon_A \| < \| A^{-1} \|^{-1}. \qquad (87.4)$$

Cuando el operador A es perturbado en ε_A, el operador inverso A^{-1} recibirá una perturbación igual a $(A + \varepsilon_A)^{-1} - A^{-1}$. Designemos mediante

$$\delta A = \frac{\| \varepsilon_A \|}{\| A \|}, \qquad \delta A^{-1} = \frac{\| (A + \varepsilon_A)^{-1} - A^{-1} \|}{\| A^{-1} \|} \qquad (87.5)$$

las magnitudes de *las perturbaciones relativas* de los operadores A y A^{-1}. Al cumplirse las condiciones (87.4), el operador $E + A^{-1}\varepsilon_A$ será regular y por eso

$$(A + \varepsilon_A)^{-1} - A^{-1} = ((A + \varepsilon_A)^{-1} A - E)\, A^{-1} =$$
$$= ((A^{-1}(A + \varepsilon_A))^{-1} - E)\, A^{-1} = ((E + A^{-1}\varepsilon_A)^{-1} - E)\, A^{-1}.$$

De acuerdo con la fórmula (87.3), para $p = 0$, encontramos que

$$\| (A + \varepsilon_A)^{-1} - A^{-1} \| \leq \frac{\| A^{-1}\varepsilon_A \| \cdot \| A^{-1} \|}{1 - \| A^{-1}\varepsilon_A \|} \leq \frac{\| A^{-1} \|^2 \, \| \varepsilon_A \|}{1 - \| A^{-1} \| \cdot \| \varepsilon_A \|}.$$

Ahora, teniendo presentes las designaciones (87.5), obtenemos la siguiente estimación:

$$\delta A^{-1} \leq \frac{\nu_A \delta A}{1 - \nu_A \delta A}, \qquad (87.6)$$

donde

$$\nu_A = \| A^{-1} \| \cdot \| A \|. \qquad (87.7)$$

El número ν_A se denomina *número de condicionalidad* del operador A. Aunque este número depende de la norma elegida, nunca puede ser muy pequeño. De la igualdad

$$E = A^{-1}A$$

concluimos, tomando en consideración (84.3), que

$$1 \leqslant \|E\| \leqslant \|A^{-1}\| \cdot \|A\| = \nu_A.$$

La fórmula (87.6) muestra que la pequeña perturbación relativa del operador A conduce a una pequeña perturbación relativa de A^{-1} sólo en el caso en que el número de condicionalidad del operador A *no es demasiado grande* en comparación con la unidad. Con este número nos encontraremos también resolviendo otros problemas.

Supongamos que con un operador regular A se resuelve la ecuación operacional

$$Ax = y. \tag{87.8}$$

Examinemos una ecuación perturbada

$$(A+\varepsilon_A)\tilde{x} = y + \varepsilon_y. \tag{87.9}$$

Si se cumple la condición (87.4), la ecuación perturbada (87.9) y la exacta (87.8) tendrán las únicas soluciones $\tilde{x}$ y x. Evaluemos su diferencia.

A la par con (87.5), (87.7) introduzcamos las designaciones correspondientes para las perturbaciones *relativas* en x, y, es decir,

$$\delta x = \frac{\|\tilde{x}-x\|}{\|x\|}, \qquad \delta_y = \frac{\|\varepsilon_y\|}{\|y\|}.$$

Tenemos

$$x = A^{-1}y, \qquad \tilde{x} = (A+\varepsilon_A)^{-1}(y+\varepsilon_y).$$

De aquí encontremos

$$\tilde{x}-x = ((E+A^{-1}\varepsilon_A)^{-1} - E)\,A^{-1}y + (E+A^{-1}\varepsilon_A)^{-1}A^{-1}\varepsilon_y$$

y luego

$$\|\tilde{x}-x\| \leqslant \|(E+A^{-1}\varepsilon_A)^{-1} - E\| \cdot \|x\| + \|(E+A^{-1}\varepsilon_A)^{-1}\| \times \\ \times \|A^{-1}\| \cdot \|\varepsilon_y\|.$$

Convengamos en considerar que se emplea la norma subordinada. Tomando en consideración las estimaciones (87.2), (87.3), como

también la desigualdad $\|y\| \leqslant \|A\| \cdot \|x\|$, obtendremos

$$\|\tilde{x}-x\| \leqslant \frac{\|A^{-1}\|\cdot\|\varepsilon_A\|\cdot\|x\|}{1-\|A^{-1}\|\cdot\|\varepsilon_A\|}+\frac{\|A^{-1}\|\cdot\|\varepsilon_y\|}{1-\|A^{-1}\|\cdot\|\varepsilon_A\|} \leqslant$$

$$\leqslant \frac{\|A^{-1}\|\cdot\|A\|\frac{\|\varepsilon_A\|}{\|A\|}}{1-\|A^{-1}\|\cdot\|A\|\frac{\|\varepsilon_A\|}{\|A\|}}\|x\|+\frac{\|A^{-1}\|\cdot\|A\|\frac{\|\varepsilon_y\|}{\|y\|}}{1-\|A^{-1}\|\cdot\|A\|\frac{\|\varepsilon_A\|}{\|A\|}}\|x\|.$$

De acuerdo con las designaciones aceptadas esto significa que

$$\delta x \leqslant \frac{\nu_A}{1-\nu_A\delta_A}(\delta A+\delta y). \qquad (87.10)$$

La fórmula obtenida muestra nuevamente la significación del número de condicionalidad y esta vez también es importante, desde el punto de vista de estabilidad, que este número sea no *demasiado grande*.

Ejercicios.

1. Demuéstrese que un número de condicionalidad expresado en forma espectral es igual a la razón entre el número singular máximo y el mínimo.
2. Existen los operadores cuyo número de condicionalidad es mínimo. ¿Qué representan en sí estos operadores, si se emplea la norma espectral?
3. Demuéstrese que si un operador se multiplica por operadores unitarios, su número de condicionalidad, expresado en la norma espectral o euclidiana, no varía.
4. Demuéstrese que para cualesquiera operadores regulares A, B se verifica la desigualdad

$$\frac{\|B^{-1}-A^{-1}\|}{\|B^{-1}\|} \leqslant \nu_A \frac{\|A-B\|}{\|A\|}.$$

5. ¿En qué radica la razón por la que el sistema de vectores descrito en el § 22 es muy inestable? Evalúese el número de condicionalidad de un operador en el que las columnas de la matriz coinciden con las coordenadas de los vectores (22.7).

§ 88. Solución estable de las ecuaciones

La fórmula (87.10) muestra que para un operador, próximo a un operador degenerado, *pueden observarse grandes perturbaciones de la solución, incluso cuando las perturbaciones en el operador y en el segundo miembro son pequeñas*. Puede parecer que este hecho sólo se debe a que la propia solución no siempre existe. Sin embargo, la situación es análoga en el caso en que se definen las seudosoluciones.

Efectivamente, supongamos que un operador actúa en un espacio bidimensional. Supongamos, además, que en cierta base ortonormalizada, a la ecuación (85.1) le corresponde un sistema de ecuaciones

algebraicas lineales del tipo siguiente:

$$1\cdot x_1 + 0\cdot x_2 = 1$$
$$0\cdot x_1 + 0\cdot x_2 = 1.$$

Es fácil determinar que la seudosolución normal u_0 tendrá las siguientes coordenadas:

$$u_0 = (1,\ 0).$$

Es muy posible que la ecuación perturbada conducirá en la misma base al sistema

$$1\cdot x_1 + 0\cdot x_2 = 1,$$
$$0\cdot x_1 + \varepsilon\cdot x_2 = 1,$$

donde el número ε, aunque pequeño, será, sin embargo distinto de cero. Ahora la seudosolución normal $u_0^{(\varepsilon)}$ de la ecuación perturbada tiene las siguientes coordenadas

$$u_0^{(\varepsilon)} = (1,\ \varepsilon^{-1}).$$

Cuando ε son pequeños, los vectores u_0 y $u_0^{(\varepsilon)}$ no sólo son muy diferentes, sino que son *casi ortogonales*.

Si una ecuación tiene más de una seudosolución, en el caso general, las perturbaciones pequeñas en el operador y en el segundo miembro siempre causarán grandes perturbaciones en la seudosolución normal. No obstante, mostraremos que *a pesar de que muchas nociones relacionadas con las ecuaciones operacionales son inestables, la seudosolución normal puede ser definida de una manera estable.*

Supongamos que el operador A actúe del espacio X en el espacio Y y que, además, se resuelve la ecuación (85.1). Por analogía con la funcional del residuo, consideraremos la así llamada *funcional de regularización*

$$\Phi_\alpha(x) = \alpha \| x \|^2 + \| Ax - y \|^2, \tag{88.1}$$

donde $\alpha \geqslant 0$. Está claro que para $\alpha = 0$ este funcional coincide con la del residuo y alcanza su mínimo en las seudosoluciones de la ecuación (85.1). Aclaremos, en qué vectores alcanza el mínimo la funcional de regularización para $\alpha > 0$. Haciendo uso de la descomposición (88.2), encontramos

$$\Phi_\alpha(x) = \sum_{k=1}^{t} (\alpha \mid \alpha_k \mid^2 + \mid \rho_k\alpha_k - \beta_k \mid^2) + \alpha \sum_{k=t+1}^{m} \mid \alpha_k \mid^2 + \sum_{p=t+1}^{n} \mid \beta_p \mid^2.$$

De aquí se deduce que para alcanzar el mínimo, es necesario tomar los valores nulos de las últimas coordenadas $\alpha_{t+1}, \ldots, \alpha_m$ y minimizar, para todo $k \leqslant t$, la expresión

$$\alpha \mid \alpha_k \mid^2 + \mid \rho_k\alpha_k - \beta_k \mid^2.$$

Esto nos da, para $k \leqslant t$,

$$\alpha_k = \frac{\rho_k\beta_k}{\alpha+\rho_k^2}.$$

De suerte que, el valor mínimo de la funcional de regularización (88.1) se alcanza para todo $\alpha > 0$ en *el único* vector

$$x_\alpha = \sum_{k=1}^{t} \frac{\rho_k\beta_k}{\alpha+\rho_k^2}\, x_k. \tag{88.2}$$

La comparación de las fórmulas (86.4), (88.2) permite establecer ciertas correlaciones que ligan x_α y x_0. Para $\rho,\ \alpha > 0$ tenemos

$$0 \leqslant \frac{|\beta|^2}{\rho^2} - \frac{\rho^2|\beta|^2}{(\alpha+\rho^2)^2} = \frac{|\beta|^2\alpha^2 + 2|\beta|^2\alpha\rho^2}{\rho^2(\alpha+\rho^2)^2} =$$

$$= \frac{2\alpha|\beta|^2\left(\frac{\alpha}{2}+\rho^2\right)}{\rho^2(\alpha+\rho^2)^2} \leqslant \frac{2\alpha|\beta|^2}{\rho^4},$$

por lo cual se desprende que

$$|x_\alpha|^2 \leqslant |x_0|^2 \leqslant |x_\alpha|^2 + 2\alpha\eta^2, \tag{88.3}$$

donde

$$\eta^2 = \sum_{k=1}^{t} \frac{|\beta_k|^2}{\rho_k^4}.$$

A continuación encontramos

$$x_0 - x_\alpha = \alpha \sum_{k=1}^{t} \frac{\beta_k}{\rho_k(\alpha+\rho_k^2)}\, x_k,$$

de donde concluimos que

$$|x_0 - x_\alpha| \leqslant \alpha\gamma, \tag{88.4}$$

donde

$$\gamma^2 = \sum_{k=1}^{t} \frac{|\beta_k|^2}{\rho_k^6}.$$

Por consiguiente,

$$\lim_{\alpha\to+0} x_\alpha = x_0.$$

De este modo, cuando los valores de α son pequeños, el vector x_α puede servir de *aproximación* a la seudosolución normal x_0.

Descompongamos los vectores x_α y x_0 según las bases singulares, por analogía con (86.2). Por comprobación directa es fácil convencerse de que x_α satisface la ecuación

$$(A^*A + \alpha E)\,x_\alpha = A^*y. \tag{88.5}$$

Para $\alpha > 0$ el operador $A^*A + \alpha E$ es definido positivo, por lo cual existe el operador $(A^*A + \alpha E)^{-1}$, es decir,

$$x_\alpha = (A^*A + \alpha E)^{-1}A^*y. \tag{88.6}$$

En el vector x_α se alcanza el valor mínimo de la funcional (88.1), por consiguiente, $\Phi_\alpha(x_\alpha) \leqslant \Phi_\alpha(x_0)$. Teniendo presentes (88.3), (88.4), obtenemos

$$|Ax_\alpha - y|^2 \leqslant |Ax_0 - y|^2 + \alpha(|x_0|^2 - |x_\alpha|^2) \leqslant \leqslant |Ax_0 - y|^2 + 2\alpha^2\eta^2. \tag{88.7}$$

Además, $\Phi_\alpha(x_\alpha) \leqslant \Phi_\alpha(0)$, de donde se infiere

$$|x_\alpha| \leqslant \frac{|y|}{\alpha^{1/2}}.$$

Junto con (88.6) esto es testimonio de que para cualquier operador A y todo vector y se tiene, para $\alpha > 0$,

$$|(A^*A + \alpha E)^{-1}A^*y| \leqslant \frac{|y|}{\alpha^{1/2}}. \tag{88.8}$$

En la resolución *práctica* de la ecuación (85.1) el operador A y el segundo miembro y se fijan, corrientemente, de manera inexacta y nos vemos obligados a considerar, en lugar de ellos, el operador perturbado $\tilde{A}$ y el segundo miembro, también perturbado, $\tilde{y}$. Si en los espacios X, Y se emplea a título de norma la longitud de los vectores, entonces a ésta última le queda subordinada la norma espectral de los operadores. Por esta razón supondremos que

$$\|A - \tilde{A}\|_2 \leqslant \tilde{\varepsilon}_A, \qquad \|y - \tilde{y}\| \leqslant \tilde{\varepsilon}_y. \tag{88.9}$$

La determinación de la solución aproximada $\tilde{x}_\alpha$, partiendo de $\tilde{A}$ e $\tilde{y}$ perturbados, conduce a una ecuación de tal índole:

$$(\tilde{A}^*\tilde{A} + \alpha E)\tilde{x}_\alpha = \tilde{A}^*\tilde{y}. \tag{88.10}$$

De (88.5), (88.10) encontramos

$$(\tilde{A}^*\tilde{A} + \alpha E)(\tilde{x}_\alpha - x_\alpha) = A^*(Ax_\alpha - y) - \tilde{A}^*(\tilde{A}x_\alpha - \tilde{y}) = = (A - \tilde{A})^*(Ax_\alpha - y) - \tilde{A}^*((\tilde{A} - A)x_\alpha - (\tilde{y} - y)).$$

Esto significa que la diferencia $\tilde{x}_\alpha - x_\alpha$ es la solución de la ecuación con el operador $(\tilde{A}^*\tilde{A} + \alpha E)$ y el segundo miembro del tipo $z = u + \tilde{A}^*v$, donde

$$u = (A - \tilde{A})^*(Ax_\alpha - y),$$
$$v = -((\tilde{A} - A)x_\alpha - (\tilde{y} - y)).$$

Por eso

$$\tilde{x}_\alpha - x_\alpha = (\tilde{A}^*\tilde{A} + \alpha E)^{-1} n + (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \tilde{A}^* v.$$

Evaluemos ahora las normas de ambos sumandos en esta igualdad.

Los valores propios del operador $\tilde{A}^*\tilde{A} + \alpha E$ no son inferiores a α. Por consiguiente, los valores propios del operador $(\tilde{A}^*\tilde{A} + \alpha E)^{-1}$ no son superiores a α^{-1}. Para un operador definido positivo la norma espectral coincide con el valor propio máximo, es decir,

$$\| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \|_2 \leqslant \alpha^{-1}.$$

Teniendo presentes (88.7), (88.9), tendremos

$$\| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} n \| \leqslant \| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \|_2 \| n \| \leqslant$$

$$\leqslant \frac{\bar{\varepsilon}_A}{\alpha} \| A x_\alpha - y \| \leqslant \frac{\bar{\varepsilon}_A}{\alpha} (\| A x_\alpha - y \|^2 + 2\alpha^2\eta^2)^{1/2}.$$

Para estimar el segundo sumando, haremos uso de las fórmulas (88.3), (88.8), (88.9). Encontramos

$$\| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \tilde{A}^* v \| \leqslant \frac{\| v \|}{\alpha^{1/2}} \leqslant \frac{1}{\alpha^{1/2}} (\bar{\varepsilon}_A \| x_0 \| + \bar{\varepsilon}_y).$$

Así pues,

$$\| \tilde{x}_\alpha - x \| \leqslant \frac{\bar{\varepsilon}_A}{\alpha} (\| A x_\alpha - y \|^2 + 2\alpha^2\eta^2)^{1/2} + \frac{1}{\alpha^{1/2}} (\bar{\varepsilon}_A \| x_0 \| + \bar{\varepsilon}_y).$$

El error total de la seudosolución calculada $\tilde{x}_\alpha$ es

$$\| x_0 - \tilde{x}_\alpha \| \leqslant \| x_0 - x_\alpha \| + \| \tilde{x}_\alpha - x_\alpha \| \leqslant$$

$$\leqslant \alpha\gamma + \frac{\bar{\varepsilon}_A}{\alpha} (\| A x_0 - y \|^2 + 2\alpha^2\eta^2)^{1/2} +$$

$$+ \frac{1}{\alpha^{1/2}} (\bar{\varepsilon}_A \| x_0 \| + \bar{\varepsilon}_y). \quad (88.11)$$

El segundo miembro de esta desigualdad no contiene ninguna información referente a los perturbados $\tilde{A}$, $\tilde{y}$ dados. Por ello existe tal α, para el cual el miembro citado alcanza su mínimo. Este valor de α asegurará la aproximación casi mejor de $\tilde{x}_\alpha$ a la seudosolución normal exacta x_0.

Supongamos que $\bar{\varepsilon}_A$ y $\bar{\varepsilon}_y$ son unas magnitudes de orden ε y el propio ε es suficientemente pequeño. Si la ecuación exacta (85.1) tiene solución, entonces $A x_0 - y = 0$. En este caso el segundo miembro de (88.11) es, a juzgar por el carácter de dependencia de α

y es una función del tipo

$$\alpha + \varepsilon + \frac{\varepsilon}{\alpha^{1/2}}.$$

Cuando $\alpha = \varepsilon^{2/3}$, esta función toma un valor de orden $\varepsilon^{2/3}$. En cambio, si la ecuación exacta no tiene ninguna solución, entonces $A x_0 - y \neq 0$. Ahora el segundo miembro de (88.11) es una función del tipo

$$\alpha + \frac{\varepsilon}{\alpha} + \frac{\varepsilon}{\alpha^{1/2}}.$$

Cuando $\alpha = \varepsilon^{1/2}$, ella toma un valor de orden $\varepsilon^{1/2}$.

De este modo, si los datos de entrada de la ecuación (85.1) *están prefijados con una exactitud del orden* ε, *la seudosolución normal puede hallarse con la exactitud del orden* $\varepsilon^{2/3}$, *si la ecuación exacta es soluble, y con una exactitud del orden* $\varepsilon^{1/2}$, *en el caso contrario.*

El parámetro α, que asegura la aproximación necesaria de $\tilde{x}_\alpha$, no puede ser determinado, partiendo sólo de $\tilde{A}$ e $\tilde{y}$ perturbados. Esto se debe, principalmente, a que las condiciones (88.9) no garantizan la continuidad de la seudosolución normal en el campo dado de variación del operador y del segundo miembro. Con el fin de determinar el parámetro α se usa, habitualmente, una información adicional acerca de la solución. En algunos problemas no se requiere una proximidad garantizada a la seudosolución normal, sino que se considera suficiente la definición estable del mínimo de la funcional del residuo. En los problemas de este tipo la determinación del parámetro resulta algo más simple. Aunque todas estas cuestiones son muy importantes, no nos detendremos ante ellas, puesto que salen de los márgenes de este curso.

Ejercicios.

1. Demuéstrese que η en la estimación (88.3) es la longitud de la solución normal de la ecuación

$$A^*A\,(A^*A)^{1/2} v = A^* y.$$

2. Demuéstrese que γ en la estimación (88.4) es la longitud de la solución normal de la ecuación

$$(A^*A)^2 v = A^* y.$$

3. Demuéstrese que la diferencia $x_\alpha - x_\beta$ satisface la ecuación

$$(A^*A + \alpha E)(A^*A + \beta E)(x_\alpha - x_\beta) = (\beta - \alpha) A^* y.$$

4. Compárense (88.11) y (87.10). ¿Qué puede decirse sobre la estimación (88.11) en el caso de un operador regular A?

5. ¿Con qué exactitud puede calcularse la seudosolución normal, si $A = 0$?

§ 89. La perturbación y los valores propios

La perturbación de un operador conduce, generalmente, a la variación de todos sus valores propios y vectores propios. Siendo muy compleja la investigación de dicha dependencia, nos limitaremos a ilustrarla con unos ejemplos. Resulta *más cómodo* describir el problema dado en términos de las matrices de los operadores en vez de considerarlo en términos de los propios operadores.

Sea B una matriz arbitraria de estructura simple y sea H una matriz tal que

$$H^{-1}BH = \Lambda, \qquad (89.1)$$

donde Λ es la matriz diagonal de los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$. Consideremos una matriz perturbada $B + \varepsilon_B$ y alguno de sus valores propios λ. La matriz $B + \varepsilon_B - \lambda E$ es degenerada, por lo cual será degenerada también la matriz

$$H^{-1}(B + \varepsilon_B - \lambda E)H = (\Lambda - \lambda E) + H^{-1}\varepsilon_B H.$$

Se presentan dos posibles casos:

1) $\lambda = \lambda_i$ para cierto i,

2) $\lambda \neq \lambda_i$ para cualquier valor de i.

En el segundo caso la matriz $\Lambda - \lambda E$ es regular, por consiguiente,

$$(\Lambda - \lambda E) + H^{-1}\varepsilon_B H = (\Lambda - \lambda E)(E + (\Lambda - \lambda E)^{-1}H^{-1}\varepsilon_B H).$$

La matriz, que interviene como segundo factor, es degenerada. Esto significa que toda norma de la matriz $(\Lambda - \lambda E)^{-1}H^{-1}\varepsilon_B H$ debe ser no inferior a la unidad. En particular,

$$\|(\Lambda - \lambda E)^{-1}H^{-1}\varepsilon_B H\|_2 \geq 1.$$

De aquí se deduce que

$$\max_{1 \leq i \leq n} |(\lambda_i - \lambda)^{-1}| \, \|H^{-1}\|_2 \|\varepsilon_B\|_2 \|H\|_2 \geq 1$$

o bien

$$\min_{1 \leq i \leq n} |\lambda_i - \lambda| \leq \|H^{-1}\|_2 \|\varepsilon_B\|_2 \|H\|_2.$$

En el primer caso esta desigualdad se verifica también, por lo cual

$$|\lambda_i - \lambda| \leq \nu_H \|\varepsilon_B\|_2 \qquad (89.2)$$

por lo menos para un valor de i. Aquí

$$\nu_H = \|H^{-1}\|_2 \|H\|_2$$

es el número de condicionalidad de la matriz H expresado por la norma espectral.

La correlación obtenida quiere decir que cualquiera que sea la perturbación ε_B de la matriz B, para todo valor propio λ de la matriz perturbada $B + \varepsilon_B$ existe tal valor propio λ_i de la matriz B que se verifique la desigualdad (89.2). Cabe notar que en nuestros razonamientos *no hemos requerido que las perturbaciones ε_B fueran pequeñas.* La correlación (89.2) puede ser interpretada de una manera algo diferente. A saber:

Los valores propios de una matriz perturbada se disponen en un dominio que representa la unión de todos los círculos de radio $\nu_H \| \varepsilon_B \|_1$ y centros en λ_i.

Las columnas de la matriz H representan en sí los vectores propios de la matriz B. Por ello, de (89.2) se infiere que como medida general de la sensibilidad de los valores propios a la perturbación de una matriz puede, evidentemente, servir el número de condicionalidad de la matriz H compuesta de los vectores propios (¡no de la matriz B!). La matriz H, que satisface (89.1), no es única, puesto que los valores propios están definidos, salvo unos factores arbitrarios. Convengamos en considerar que la matriz se elige siempre de tal manera que el valor ν_H quede mínimo. Recordemos que en todo caso $\nu_H \geqslant 1$.

Si B es una matriz normal y, además, hermitiana o unitaria, entonces H siempre puede elegirse unitaria. En este caso $\nu_H = 1$ y, por lo tanto,

$$|\lambda_i - \lambda| \leqslant \| \varepsilon_B \|_1. \tag{89.3}$$

Examinemos más detalladamente el caso de una matriz hermitiana B con la perturbación hermitiana ε_B. Ahora podemos mostrar que

En todo círculo con el centro en λ_i, de radio $\| \varepsilon_B \|_1$ está contenido por lo menos un solo valor propio de la matriz perturbada.

Efectivamente, consideraremos *convencionalmente* la matriz $B + \varepsilon_B$ como «exacta» y la matriz $B = (B + \varepsilon_B) - \varepsilon_B$, como «perturbada» cuya perturbación es igual a $-\varepsilon_B$. Repitiendo textualmente todos los razonamientos, obtendremos una fórmula análoga a la (89.3) con la particularidad de que en la primera los valores propios de las matrices B y $B + \varepsilon_B$ cambian sus papeles. Esto significa que para todo valor propio λ_i de la matriz «perturbada» B existe indefectiblemente por lo menos un solo valor propio λ de la matriz «exacta» $B + \varepsilon_B$, para el cual la desigualdad (89.3) tiene lugar.

Si los valores propios de la matriz B son simples, entonces siendo la perturbación ε_B suficientemente pequeña, todos los círculos *se separan* y todo círculo contendrá *uno*, y *sólo* un valor propio de la matriz perturbada.

La fórmula (89.3) muestra que los valores propios de las matrices normales poseen una estabilidad considerable a la perturbación. No obstante, en el problema general de la definición de los valores propios este fenómeno es más bien una excepción que una regla,

A título de ejemplo consideraremos un caso, «límite» en cierto sentido, en que la matriz B se compone de una sola caja canónica de Jordan. Se puede considerar convencionalmente que todos los vectores propios de tal matriz son colineales, la matriz de vectores propios es degenerada y, por consiguiente, su número de condicionalidad es igual a «infinito». Así pues, supongamos que la matriz B de orden m tiene por expresión

$$B = \begin{pmatrix} \lambda_0 & 1 & & & & 0 \\ & \lambda_0 & 1 & & & \\ & & \lambda_0 & 1 & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \\ 0 & & & & \lambda_0 & 1 \\ & & & & & \lambda_0 \end{pmatrix}.$$

Es evidente que su polinomio característico es $(\lambda - \lambda_0)^m$.

Tomemos ahora tal matriz de perturbación ε_B en la cual sólo un elemento, dispuesto en posición $(m, 1)$ es diferente de cero y es igual al número ε. El polinomio característico de la matriz perturbada es igual a $(\lambda - \lambda_0)^m - \varepsilon$. Por ello, los valores propios de la matriz perturbada se encuentran a la distancia de $|\varepsilon|^{1/m}$ de los valores propios de la matriz exacta. Si, por ejemplo, $m = 20$, $\varepsilon = 10^{-10}$ y λ_0 es de orden uno, entonces no se puede hablar de ninguna estabilidad práctica.

Es importante comprender que la inestabilidad de los valores propios no está ligada obligatoriamente a la presencia de valores propios múltiples y tampoco, con la mayor razón, a la presencia de las cajas de Jordan. Examinemos una matriz de orden 20

$$B = \begin{pmatrix} 20 & 20 & & & & 0 \\ & 19 & 20 & & & \\ & & 18 & 20 & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \\ 0 & & & & 2 & 20 \\ & & & & & 1 \end{pmatrix}.$$

Es una matriz triangular y, por eso, sus valores propios se constituirán de elementos diagonales. A primera vista, están suficientemente bien separados y parecería que no hay ninguna razón de esperar la inestabilidad. Pero, agreguemos la perturbación ε al elemento nulo dispuesto en la posición (20,1). El término independiente del polinomio característico variará en este caso a una magnitud de $20^{19}\varepsilon$.

Dado que el producto de los valores propios es igual al término independiente, los mismos valores propios deben alterarse en un grado considerable.

Unos problemas, más complejos aún, surgen al estudiar la estabilidad de los vectores propios. Está claro que si un valor propio λ de la matriz B es inestable ante una perturbación, entonces el vector propio x, que corresponde a dicho valor, no puede ser estable a ciencia cierta, puesto que B, λ, x están ligados entre sí mediante la correlación lineal $Bx = \lambda x$.

Sin embargo, resulta importante subrayar que si incluso los valores propios no varían como resultado de una perturbación, los vectores propios no sólo pueden ser inestables, sino que su número puede variar. Por ejemplo, la primera de las matrices

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & \varepsilon \\ 0 & 0 & 1 \end{pmatrix}$$

tiene tres vectores propios linealmente independientes y la segunda, dos vectores, aunque los valores propios de estas matrices son iguales. Desde el punto de vista teórico, este fenómeno está relacionado sólo con el hecho de presencia de valores propios múltiples de la matriz inicial. Pero, *cuando una matriz se da aproximadamente, es difícil y frecuentemente imposible decir cuáles valores propios deben considerarse múltiples y cuáles, simples.*

Los problemas referentes al estudio de la estabilidad de los valores propios, los vectores propios y radicales aparecen como los más complejos en los apartados del álgebra relacionados con los cálculos.

Ejercicios.

1. Supongamos que una matriz B es de estructura simple, pero tiene valores propios múltiples. Demuéstrese que para cualquier número $\varepsilon > 0$, tan pequeño como se quiera, existe tal perturbación ε_B, la cual satisface la condición $\|\varepsilon_B\| < \varepsilon$, que la matriz $B + \varepsilon_B$ ya no será de estructura simple.

2. Supongamos que una matriz B tiene valores propios distintos dos a dos y que $d > 0$ es la distancia mínima entre los valores propios. Demuéstrese que existe tal perturbación ε_B, la cual satisface la condición $\|\varepsilon_B\|_1 > d$, que la matriz $B + \varepsilon_B$ no será de estructura simple.

3. Supongamos ahora que la matriz B es hermitiana. Demuéstrese que si una perturbación hermitiana ε_B satisface la condición $\|\varepsilon_B\|_1 < d/2$, la matriz $B + \varepsilon_B$ tiene valores propios distintos dos a dos.

4. Supongamos, por fin, que la matriz B no es hermitiana y tiene valores propios distintos dos a dos. Demuéstrese que existe tal número r, el cual satisface la condición $0 < r \leq d$, que la matriz $B + \varepsilon_B$ tiene estructura simple, siempre que $\|\varepsilon_B\|_2 \leq r$.

5. Trátese de establecer una relación más estrecha entre los números r y d.

PARTE III FORMAS BILINEALES

CAPÍTULO 11 FORMAS BILINEALES Y CUADRÁTICAS

§ 90. Propiedades generales de las formas bilineales y cuadráticas

Veamos las funciones numéricas $\varphi(x, y)$ de dos argumentos vectoriales x, y de cierto espacio lineal K_n, dado sobre un campo numérico P, x, y tomen los valores de P. Una función $\varphi(x, y)$ se llama *forma bilineal*, si para cualesquiera dos vectores x, y, $z \in K_n$ y todo número $\alpha \in P$ se verifican las correlaciones

$$\varphi(x + z, y) = \varphi(x, y) + \varphi(z, y), \quad \varphi(\alpha x, y) = \alpha\varphi(x, y),$$
$$\varphi(x, y + z) = \varphi(x, y) + \varphi(x, z), \quad \varphi(x, \alpha y) = \alpha\varphi(x, y). \qquad (90.1)$$

Las primeras dos correlaciones de (90.1) significan la linealidad de la forma $\varphi(x, y)$ respecto del primer argumento, las dos últimas, la linealidad respecto del segundo argumento.

Es fácil comprobar que la suma de dos formas bilineales, como también el producto de una forma bilineal por un número será nuevamente una forma bilineal. Por esta razón, el conjunto de todas las formas bilineales, definidas sobre un mismo espacio K_n, que toman los valores de un mismo campo numérico P será un espacio lineal. En este caso el cero del espacio dado será la forma bilineal $0(x, y)$, para la cual $0(x, y) = 0$, cualesquiera que sean x, y. $0(x, y)$ se denomina forma bilineal *nula*.

Anteriormente ya nos encontramos con una función de tal índole. Comparando (27.1) y (90.1), se nota con facilidad que un producto escalar en un espacio euclídeo es una forma bilineal. Recordando el papel importante que desempeñó el producto escalar al estudiar los espacios euclídeos y los operadores lineales que actúan en los mismos, podemos suponer que el estudio de las formas bilineales será también útil.

Entre las formas bilineales las simétricas y las antisimétricas atraen un interés singular. Una forma bilineal $\varphi(x, y)$ se llama *simétrica*, si para cualesquiera vectores $x, y \in K_n$ se verifica la desigualdad

$$\varphi(x, y) = \varphi(y, x).$$

En cambio, si para cualesquiera $x, y \in K_n$

$$\varphi(x, y) = -\varphi(y, x),$$

la forma bilineal se denomina *antisimétrica*.

Toda forma bilineal antisimétrica $\varphi(x, y)$ se anula, si los argumentos coinciden. En efecto, como $\varphi(x, x) = -\varphi(x, x)$, entonces $\varphi(x, x) = 0$. Algo sorprendente parece el otro hecho, vinculado con los valores de una forma bilineal simétrica al coincidir los argumentos. A saber, cada forma bilineal simétrica $\varphi(x, y)$ se define unívocamente por sus valores, siendo coincidentes los argumentos. Efectivamente, sean x, y cualesquiera vectores de K_n. Tomando en consideración la simetría de la forma $\varphi(x, y)$, tenemos

$$\varphi(x + y, x + y) = \varphi(x, x) + \varphi(y, y) + 2\varphi(x, y), \qquad (90.2)$$

de donde se desprende que

$$\varphi(x, y) = \frac{1}{2}\{\varphi(x + y, x + y) - \varphi(x, x) - \varphi(y, y)\}. \qquad (90.3)$$

La fórmula obtenida demuestra que la afirmación enunciada es justa, puesto que el segundo miembro de la correlación es una forma bilineal simétrica.

Una forma bilineal se descompone unívocamente en la suma de las formas bilineales simétrica y antisimétrica. Esta descomposición puede escribirse explícitamente

$$\varphi(x, y) = \frac{1}{2}\{\varphi(x, y) - \varphi(y, x)\} + \frac{1}{2}\{\varphi(x, y) - \varphi(y, x)\}. \qquad (90.4)$$

Es fácil comprobar que los primeros dos sumandos en el miembro derecho dan una forma bilineal simétrica y los dos últimos, antisimétrica. Si se admite la existencia de alguna otra descomposición, entonces, al sustituir los argumentos iguales, habremos de hacer una deducción sobre la univocidad de la definición de la parte simétrica de la descomposición y, consecuentemente, de la descomposición en total.

Si una forma bilineal no es simétrica, en lugar de (90.2) tendremos

$$\varphi(x + y, x + y) = \varphi(x, x) + \varphi(y, y) + \varphi(x, y) + \varphi(y, x).$$

Por consiguiente,

$$\frac{1}{2}\{\varphi(x, y) + \varphi(y, x)\} = \frac{1}{2}\{\varphi(x + y, x + y) - \varphi(x, x) - \varphi(y, y)\}. \qquad (90.5)$$

Comparando la correlación obtenida con (90.3), concluimos que para la forma bilineal antisimétrica su parte simétrica se define unívocamente por los valores de la forma, siendo coincidentes los argumentos.

A la par con las formas bilineales consideraremos también las así llamadas formas cuadráticas. Sea $\varphi(x, y)$ una forma bilineal en el espacio K_n. Se denomina *forma cuadrática* la función numérica $\varphi(x, x)$ de un solo argumento vectorial $x \in K_n$, la cual se obtiene de la forma bilineal $\varphi(x, y)$ sustituyendo el vector y por el vector x.

En general, partiendo de la forma cuadrática, no se puede restablecer unívocamente la forma bilineal que la ha engendrado. Pero, según se deduce de la fórmula (90.3), existe una forma bilineal simétrica y sólo una de la cual puede ser obtenida la forma cuadrática inicial. Esta forma bilineal se llama *polar* respecto a la forma cuadrática dada. El conjunto de todas las formas bilineales que engendran una misma forma cuadrática puede ser obtenido por sumación de la forma bilineal polar con una forma antisimétrica arbitraria. Es por eso que al utilizar formas bilineales para estudiar las propiedades de las formas cuadráticas resulta suficiente limitarse sólo a la consideración de las formas bilineales simétricas.

El hecho de que una forma bilineal no pueda ser restablecida según la cuadrática se debe a que la última no proporciona ninguna información sobre la parte antisimétrica, cualquiera que sea la forma bilineal.

LEMA 90.1. *Las formas bilineales antisimétricas, y sólo ellas, se anulan, cuando todos los argumentos coinciden.*

DEMOSTRACIÓN. Ya hemos señalado que si $\varphi(x, y)$ es una forma antisimétrica, entonces $\varphi(x, x) = 0$ para cualquier x. En cambio, si $\varphi(x, x) = 0$ para todo x, entonces de la correlación (90.5) se infiere que para todos los vectores x, y se verifica la igualdad $\varphi(x, y) + \varphi(y, x) = 0$, es decir, la forma bilineal $\varphi(x, y)$ es antisimétrica.

La comparación de las propiedades que poseen el producto escalar y las correlaciones (90.1) muestra que en un espacio unitario el producto escalar no es, estrictamente, una forma bilineal. En un espacio complejo las formas bilineales hermitianas están estrechamente vinculadas con el producto escalar. La función numérica $\varphi(x, y)$ se llama forma bilineal *hermitiana*, si para cualesquiera vectores x, y, $z \in K_n$ y todo número α del campo de números complejos P se verifican las correlaciones

$$\varphi(x + z, y) = \varphi(x, y) + \varphi(z, y), \qquad \varphi(\alpha x, y) = \alpha\varphi(x, y),$$
$$\varphi(x, y + z) = \varphi(x, y) + \varphi(x, z), \qquad \varphi(x, \alpha y) = \bar{\alpha}\varphi(x, y).$$

La raya significa aquí una conjugación compleja.

En este caso también la suma de dos formas bilineales hermitianas y asimismo el producto de una forma bilineal hermitiana por un

número será una forma bilineal hermitiana. Por esto el conjunto de todas las formas bilineales hermitianas que están definidas sobre un espacio complejo y que toman valores complejos es un espacio lineal complejo.

La forma bilineal hermitiana se denomina *simétrica hermitiana*, si para cualesquiera vectores $x, y \in K_n$ se tiene

$$\varphi(x, y) = \overline{\varphi(y, x)}.$$

Si para cualesquiera $x, y \in K_n$

$$\varphi(x, y) = -\overline{\varphi(y, x)},$$

la forma se llama *antisimétrica hermitiana*. En los vectores coincidentes la forma antisimétrica hermitiana toma valores imaginarios puros, mientras que la forma simétrica hermitiana, valores reales. Ahora, toda forma bilineal hermitiana se define unívocamente por sus valores si los argumentos son coincidentes. Pero, en vez de (90.3), es lícita la correlación

$$\varphi(x, y) = \frac{1}{4}\{\varphi(x+y, x+y) - \varphi(x-y, x-y) + i\varphi(x+iy, x+iy) - i\varphi(x-iy, x-iy)\} \qquad (90.6)$$

De esta correlación se infiere, en particular, que

Entre las formas bilineales hermitianas la forma nula, y sólo ella, toma los valores nulos cuando todos los argumentos coinciden.

En este caso también la forma bilineal hermitiana puede ser unívocamente representada como una suma de la forma simétrica hermitiana y la antisimétrica hermitiana, con la particularidad de que

$$\varphi(x, y) = \frac{1}{2}\{\varphi(x, y) + \overline{\varphi(y, x)}\} + \frac{1}{2}\{\varphi(x, y) - \overline{\varphi(y, x)}\} \qquad (90.7)$$

Las demostraciones de los hechos enunciados para las formas hermitianas casi no se diferencian de las demostraciones correspondientes para las formas bilineales.

Se llama forma cuadrática *hermitiana* la función numérica $\varphi(x, x)$ de un solo argumento vectorial $x \in K_n$, la cual se obtiene de la función bilineal hermitiana $\varphi(x, y)$ al sustituir el vector y por el vector x. A diferencia de las formas cuadráticas, a base de una forma cuadrática hermitiana se restablece unívocamente la forma bilineal hermitiana que engendra la citada forma cuadrática hermitiana. Este restablecimiento se realiza de acuerdo con la fórmula (90.6) y la correspondiente forma bilineal se llama también *polar* respecto a la forma cuadrática de partida.

La posibilidad de restablecer unívocamente una forma bilineal hermitiana sobre la base de la forma cuadrática hermitiana es-

gendrada por la primera se explica por una estrecha relación que existe entre las formas bilineales simétricas hermitianas y antisimétricas hermitianas.

LEMA 40.2. *Si $\varphi(x, y)$ es una forma bilineal simétrica (antisimétrica) hermitiana, entonces $\psi(x, y) = i\varphi(x, y)$ será una forma bilineal antisimétrica (simétrica) hermitiana.*

DEMOSTRACIÓN. Sea, por ejemplo, $\varphi(x, y)$ una forma simétrica hermitiana. Entonces, para todos los vectores x, y se tiene

$$\psi(x, y) = i\varphi(x, y) = \varphi(ix, y) = \overline{\varphi(y, ix)} = -\overline{i\varphi(y, x)} = -\overline{\psi(y, x)},$$

es decir, $\psi(x, y)$ es una forma antisimétrica hermitiana. El caso de la forma antisimétrica hermitiana $\varphi(x, y)$ se considera de modo análogo.

En lo que sigue trataremos frecuentemente las formas cuadráticas hermitianas engendradas por las formas bilineales simétricas hermitianas.

LEMA 40.3. *Entre las formas bilineales hermitianas las simétricas, y sólo ellas, engendran formas cuadráticas hermitianas reales.*

DEMOSTRACIÓN. Anteriormente ya se ha observado que las formas simétricas hermitianas toman valores reales cuando coinciden los argumentos. Supondremos ahora que una forma cuadrática hermitiana $\varphi(x, x)$ toma sólo los valores reales. Conforme a (40.6), para la forma bilineal polar $\varphi(x, y)$ tenemos

$$\varphi(y, x) = \frac{1}{4}(\varphi(y + x, y + x) - \varphi(y - x, y - x) +$$
$$+ i\varphi(y + ix, y + ix) - i\varphi(y - ix, y - ix)) =$$
$$= \frac{1}{4}(\varphi(x + y, x + y) - \varphi(x - y, x - y) + i\varphi(x - iy, x - iy) -$$
$$- i\varphi(x + iy, x + iy)) = \frac{1}{4}(\overline{\varphi(x + y, x + y)} - \overline{\varphi(x - y, x - y)} +$$
$$+ \overline{i\varphi(x + iy, x + iy)} - \overline{i\varphi(x - iy, x - iy)}) = \overline{\varphi(x, y)}.$$

COROLARIO. *Entre las formas bilineales hermitianas las antisimétricas, y sólo ellas, engendran formas cuadráticas hermitianas imaginarias puras.*

COROLARIO. *Ninguna forma bilineal antisimétrica hermitiana puede engendrar una forma cuadrática hermitiana real.*

Según se deduce de las propiedades de linealidad de las formas bilineales y bilineales hermitianas respecto de cada argumento, $\varphi(0, 0) = 0$ para cualquier forma cuadrática $\varphi(x, x)$. Sin embargo, en el caso general pueden existir también los vectores no nulos x, para los cuales $\varphi(x, x) = 0$. Tales vectores se llamarán *isótropos*. El concepto de isotropía está ligado sólo con la forma cuadrática. Por ello, unos vectores, isótropos para una forma cuadrática, pueden

no sólo para otra forma cuadrática, y viceversa. En particular, el lema 90.1 significa que para una forma cuadrática engendrada por una forma bilineal antisimétrica, todos los vectores del espacio K_n, salvo el nulo, son isótropos.

De las formas reales hermitianas y ordinarias son de mayor uso aquellas que toman los valores de un mismo signo para todos los segmentos vectoriales. La forma cuadrática real $\varphi(x, x)$ se denomina *definida positiva*, si $\varphi(x, x) > 0$ para todo $x \neq 0$. La forma se llama *no negativa*, si para todo $x \neq 0$ se cumple la desigualdad $\varphi(x, x) \geqslant 0$. Análogamente se definen las formas cuadráticas *no positivas* y *definidas negativas*.

Como regla, solamente las formas cuadráticas definidas positivas y definidas negativas se llaman *de signo constante*. Pero, a veces, así se llaman también las formas cuadráticas no negativas y no positivas. Con el fin de evitar toda clase de equivocaciones, las formas cuadráticas definidas positivas y definidas negativas se llamarán, cuando sea necesario, *estrictamente* de signo constante.

Si una forma cuadrática real es de signo constante, la forma bilineal hermitiana u ordinaria que la engendra se llamará también definida positiva, no negativa, etc.

Si una forma cuadrática real $\varphi(x, x)$ es estrictamente de signo constante, no tiene vectores isótropos. En el caso de las formas bilineales reales y bilineales simétricas hermitianas $\varphi(x, y)$, las correspondientes formas cuadráticas serán reales y para ellas resulta cierta la afirmación recíproca. A saber, tiene lugar el

TEOREMA 90.1. *Supongamos que una forma cuadrática $\varphi(x, x)$ está engendrada por la forma bilineal real o bilineal simétrica hermitiana $\varphi(x, y)$. Si $\varphi(x, x)$ no tiene vectores isótropos, es estrictamente de signo constante.*

DEMOSTRACIÓN. Como ya se ha indicado, la forma cuadrática $\varphi(x, x)$ es real. En ambos casos toma en vectores colineales los valores de un mismo signo. Supongamos que $\varphi(x, x)$ no es estrictamente de signo constante. En este caso existen tales vectores linealmente independientes u, v que $\varphi(u, u) > 0$ y $\varphi(v, v) < 0$. Para todo número real α

$$\varphi(u + \alpha v, u + \alpha v) = \varphi(u, u) + \alpha(\varphi(u, v) + \varphi(v, u)) + \alpha^2\varphi(v, v) \qquad (90.8)$$

El segundo miembro de esta igualdad es un polinomio de segundo grado respecto de α. Sus coeficientes son reales, lo que se determina por el carácter real de la forma cuadrática $\varphi(x, x)$ y el lema 90.3. Puesto que $\varphi(u, u)$ y $\varphi(v, v)$ son de signos opuestos, el polinomio (90.8) tendrá dos raíces reales. Sea α_0 una de ellas. Esto significa que $\varphi(u + \alpha_0 v, u + \alpha_0 v) = 0$. Sin embargo, el vector $u + \alpha_0 v$ es no nulo por ser los vectores u, v linealmente independientes, por lo cual, la anulación en él de la forma cuadrática no es posible por

hipótesis del teorema. La contradicción obtenida da por terminada la demostración del mismo.

No es casual que en el teorema 90.1 nos limitamos a la consideración de las formas cuadráticas engendradas sólo por las formas bilineal real y bilineal simétrica hermitiana. Ninguna otra forma bilineal puede conducir a una forma cuadrática real. Nos queda considerar, de hecho, sólo una forma bilineal en un espacio complejo. Pero tal forma bilineal no puede engendrar una forma cuadrática real que no sea idénticamente igual a cero. Si para cierto vector u una forma cuadrática toma el valor real $\varphi(u, u)$ que no es igual a cero, entonces $\varphi(\alpha u, \alpha u) = \alpha^2 \varphi(u, u)$ será un número complejo, cualquiera que sea α complejo con las partes imaginaria pura y real no nulas. Así pues,

Para las formas cuadráticas reales la condición necesaria y suficiente de que estas formas no tengan vectores isótropos consiste en el mantenimiento estricto de signo constante.

La forma bilineal compleja genera siempre una forma cuadrática, que tiene vectores isótropos, siempre que esté definida en un espacio lineal cuya dimensión es superior a uno. En efecto, si suponemos que esto no es así, se hallarán siempre unos vectores linealmente independientes u, v, para los cuales $\varphi(u, u) \neq 0$, $\varphi(v, v) \neq 0$. Pero, de conformidad con (90.8), el vector $u + \alpha v$ será isótropo, elegido de manera adecuada el número complejo α. La forma compleja bilineal hermitiana puede generar una forma cuadrática que no tiene vectores isótropos. Según se deduce de nuestras investigaciones,

Para que una forma cuadrática engendrada por la forma bilineal hermitiana no tenga vectores isótropos, es suficiente que la parte real (o imaginaria) de la forma cuadrática sea estrictamente de signo constante.

Ejercicios.

1. Demuéstrese que para toda forma bilineal $\varphi(x, y)$ se verifican las desigualdades $\varphi(0, y) = \varphi(x, 0) = 0$, cualesquiera que sean $x, y \in K_n$.
2. Determínese la dimensión y la base de un espacio lineal de formas bilineales.
3. Demuéstrese que los conjuntos de formas bilineales simétricas y antisimétricas forman subespacios en un espacio lineal de todas las formas bilineales.
4. Demuéstrese que un espacio de todas las formas bilineales es la suma directa de los subespacios de las formas bilineales simétricas y antisimétricas.
5. Demuéstrese que un conjunto de todas las formas cuadráticas forma un espacio lineal. Determínese su dimensión y la base.
6. ¿Formarán los subespacios lineales los siguientes conjuntos de formas cuadráticas:

formas cuadráticas de signo constante,
formas cuadráticas que toman los valores reales,
formas cuadráticas que no tienen vectores isótropos,
formas cuadráticas, para las cuales todos los vectores del conjunto dado son isótropos?

7. Demuéstrese que para toda forma cuadrática, dada en un espacio normalizado, existe tal número α que para todo x

$$|\varphi(x, x)| \leqslant \alpha \|x\|^2.$$

8. Supongamos que la forma cuadrática $\varphi(x, x)$ es estrictamente de signo constante y la forma cuadrática $\psi(x, x)$ es arbitraria. Demuéstrese que existe tal número β, que para todo x

$$|\psi(x, x)| \leqslant \beta\varphi(x, x).$$

9. Demuéstrese que una forma cuadrática no es estrictamente de signo constante cuando, y sólo cuando, el conjunto de vectores isótropos y el vector nulo forman un subespacio lineal.

10. Examínense los ejercicios 1—9 para las formas bilineales hermíticas y las cuadráticas. ¿Serán verídicas todas las afirmaciones enunciadas?

11. Supongamos que en el espacio complejo K_n un subespacio L sólo consiste de los vectores isótropos de la forma bilineal hermítica $\varphi(x, y)$ y el vector nulo. Demuéstrese que $\varphi(u, v) = 0$ para cualesquiera vectores $u, v \in L$.

§ 91. Matrices de las formas bilineales y cuadráticas

Investiguemos una forma bilineal $\varphi(x, y)$ definida en el espacio K_n. Elijamos en K_n dos bases fijadas $e_1, e_2, \ldots, e_n$ y $q_1, q_2, \ldots, q_n$ y sean

$$x = \sum_{i=1}^{n} \xi_i e_i, \qquad y = \sum_{i=1}^{n} \eta_i q_i.$$

En virtud de las propiedades (90.1) tenemos

$$\varphi(x, y) = \varphi\Big(\sum_{i=1}^{n} \xi_i e_i, \sum_{j=1}^{n} \eta_j q_j\Big) = \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi(e_i, q_j)\, \xi_i \eta_j. \quad (91.1)$$

Designemos, como antes, con x_e e y_q las matrices de dimensiones $n \times 1$, compuestas por las coordenadas de los vectores x y y en las bases correspondientes, y mediante G_{eq} la matriz de orden n con los elementos $g_{ij}^{(eq)} = \varphi(e_i, q_j)$. La correlación (91.1) significa que

$$\varphi(x, y) = x_e' G_{eq} y_q. \quad (91.2)$$

De este modo, siendo fijadas las bases en el espacio K_n, la forma bilineal puede ser representada en la forma matricial (91.2).

G_{eq} se llama *matriz de la forma bilineal* y con las bases fijadas se define unívocamente. Si suponemos que para la forma $\varphi(x, y)$ existe, además de (91.2), otra representación análoga con cierta matriz F_{eq}, entonces, al poner $x = e_i$, $y = q_j$ obtenemos en seguida que $f_{ij}^{(eq)} = \varphi(e_i, q_j)$, es decir, $F_{eq} = G_{eq}$.

Cabe señalar que el segundo miembro de (91.2) define, para cualquier matriz G_{eq}, cierta forma bilineal. El cumplimiento de las correlaciones (90.1) se infiere directamente de las propiedades co-

correspondientes de las operaciones matriciales. Así se establece, con las bases fijadas en K_n, una correspondencia biunívoca entre las formas bilineales y las matrices cuadradas.

Al cambiar las bases en K_n, la matriz de la forma bilineal, desde luego, varía. Sea P una matriz de la transformación de coordenadas al pasar de la base $e_1, e_2, \ldots, e_n$ a la $f_1, f_2, \ldots, f_n$ y sea Q la matriz de la transformación de coordenadas al pasar de $q_1, q_2, \ldots, q_n$ a $t_1, t_2, \ldots, t_n$. De acuerdo con (63.3)

$$x_e = P x_f, \qquad y_q = Q y_t. \tag{91.3}$$

por lo cual de (91.2) se desprende que

$$\varphi(x, y) = x'_e G_{eq} y_q = x'_f P' G_{eq} Q y_t.$$

Pero, por otra parte,

$$\varphi(x, y) = x'_f G_{ft} y_t.$$

Por consiguiente,

$$G_{ft} = P' G_{eq} Q. \tag{91.4}$$

Como las matrices P y Q son regulares, entonces, de acuerdo con la terminología introducida en el § 64, las matrices G_{ft} y G_{eq} se llamarán equivalentes. Según se ha señalado anteriormente, las matrices equivalentes de un mismo orden, y sólo ellas, tienen rangos iguales. Esto es testimonio de que el rango de una matriz de la forma bilineal no depende de las bases elegidas y es una característica de la misma forma. Llamémoslo *rango* de la forma bilineal. Una forma bilineal se llamará *regular*, si es regular su matriz. Como característica de una forma bilineal sirve también la diferencia entre la dimensión del espacio K_n y el rango de la forma. Se llamará dicha característica *defecto* de la forma bilineal.

De los resultados del § 64 se desprende que todas las matrices de un mismo rango son equivalentes a la matriz diagonal con los elementos 0 y 1. En el lenguaje de las formas bilineales este hecho atestigua que para una forma arbitraria de rango r siempre pueden indicarse tales bases $f_1, f_2, \ldots, f_n$ y $t_1, t_2, \ldots, t_n$, en las cuales la forma tendrá una expresión más simple. A saber, si

$$x = \sum_{i=1}^{n} \tau_i f_i, \qquad y = \sum_{i=1}^{n} \nu_i t_i,$$

entonces

$$\varphi(x, y) = \sum_{i=1}^{r} \tau_i \nu_i.$$

La elección separada de las bases para toda forma bilineal variable se realiza raramente. Con mucha más frecuencia se utiliza

una base común. Sea $e_1, e_2, \ldots, e_n$ cierta base de K_n y

$$x=\sum_{i=1}^{n} \xi_i e_i, \qquad y=\sum_{j=1}^{n} \eta_j e_j.$$

En este caso, por analogía con (91.1), obtenemos la siguiente representación de la forma bilineal:

$$\varphi(x, y)=\sum_{i=1}^{n} \sum_{j=1}^{n} \varphi(e_i, e_j)\, \xi_i \eta_j$$

o, en la escritura matricial,

$$\varphi(x, y)=x_e' G_e y_e. \tag{91.5}$$

Aquí G_e es una matriz con los elementos $g_{ij}^{(e)}=\varphi(e_i e_j)$. En lo sucesivo la matriz G_e se llamará siempre matriz de la forma bilineal. Si, de nuevo, P es una matriz de la transformación de coordenadas al pasar de la base $e_1, e_2, \ldots, e_n$ a la $f_1, f_2, \ldots, f_n$, entonces las matrices G_e y G_f de la misma forma bilineal $\varphi(x, y)$ estarán ligadas entre sí, de acuerdo con (91.4), por la correlación

$$G_f=P' G_e P. \tag{91.6}$$

Las matrices G_e y G_f, ligadas mediante la correlación (91.6), siendo regular la matriz P, se denominan *congruentes*. Las matrices congruentes son siempre equivalentes. Lo recíproco, desde luego, en el caso general no es cierto.

Todo lo dicho acerca de las formas bilineales puede ser aplicado con cambios insignificantes a las formas bilineales hermitianas. Toda forma hermitiana se representa unívocamente en forma matricial

$$\varphi(x, y)=x_e' G_{eq} \bar{y}_q,$$

siendo fijadas las bases $e_1, e_2, \ldots, e_n$ y $q_1, q_2, \ldots, q_n$. Al pasar a las otras bases $f_1, f_2, \ldots, f_n$ y $t_1, t_2, \ldots, t_n$ en lugar de (91.4) tendremos

$$G_{ft}=P' G_{eq} \bar{Q}.$$

Si los argumentos de la forma bilineal hermitiana están dados en una misma base, la anotación matricial de la forma es análoga a (91.6). A saber

$$\varphi(x, y)=x_e' G_e \bar{y}_e \tag{91.7}$$

Al pasar a la nueva base, las matrices de la forma quedarán ligadas entre sí por la correlación

$$G_f=P' G_e \bar{P}$$

y se dirá que estas matrices son *congruentes según Hermite*.

Ahora se puede establecer la relación entre la expresión de una forma bilineal y la de su matriz. Si la forma es simétrica, para cualquier base $e_1, e_2, \ldots, e_n$ se tiene

$$g_{ij}^{(e)} = \varphi(e_i, e_j) = \varphi(e_j, e_i) = g_{ji}^{(e)},$$

donde $G_e = G_e'$ y la matriz G_e de la forma $\varphi(x, y)$ es *simétrica*. En cambio, si la forma es antisimétrica, entonces

$$g_{ij}^{(e)} = \varphi(e_i, e_j) = -\varphi(e_j, e_i) = -g_{ji}^{(e)},$$

es decir, $G_e = -G_e'$. En este caso la matriz G_e es también *antisimétrica*.

La afirmación recíproca es asimismo cierta. Si en una base la matriz de la forma es simétrica (antisimétrica), la forma bilineal que la genera será también simétrica (antisimétrica). Sea $G_e = G_e'$, entonces

$$\varphi(y, x) = y_e' G_e x_e = (y_e' G_e x_e)' = x_e' G_e' y_e = x_e' G_e y_e = \varphi(x, y).$$

Si, en cambio, $G_e' = -G_e$, entonces

$$\varphi(y, x) = y_e' G_e x_e = (y_e' G_e x_e)' = x_e' G_e' y_e = -x_e' G_e y_e = -\varphi(x, y).$$

Las afirmaciones análogas subsisten también respecto de la relación existente entre la forma bilineal hermitiana y su matriz. Si la forma hermitiana es simétrica, entonces

$$g_{ij}^{(e)} = \varphi(e_i, e_j) = \overline{\varphi(e_j, e_i)} = \bar{g}_{ji}^{(e)},$$

es decir, $G_e = \bar{G}_e'$ y la matriz G_e de la forma $\varphi(x, y)$ es *hermitiana*. Si la forma es antisimétrica hermitiana, entonces

$$g_{ij}^{(e)} = \varphi(e_i, e_j) = -\overline{\varphi(e_j, e_i)} = -\bar{g}_{ji}^{(e)},$$

donde $G_e = -\bar{G}_e'$. En este caso la matriz G_e se denomina *antihermitiana*.

Son ciertas también las afirmaciones recíprocas. Sea $G_e = \bar{G}_e'$, entonces para una forma bilineal hermitiana generadora tenemos

$$\varphi(y, x) = y_e' G_e \bar{x}_e = (y_e' G_e \bar{x}_e)' = \bar{x}_e' G_e' y_e = \overline{x_e' \bar{G}_e' \bar{y}_e} = \overline{x_e' G_e \bar{y}_e} = \overline{\varphi(x, y)}.$$

Para el caso en que $G_e = -\bar{G}_e'$ encontramos

$$\varphi(y, x) = y_e' G_e \bar{x}_e = (y_e' G_e \bar{x}_e)' = \bar{x}_e' G_e' y_e = \overline{x_e' \bar{G}_e' \bar{y}_e} = -\overline{x_e' G_e \bar{y}_e} = -\overline{\varphi(x, y)}.$$

La matriz de la forma bilineal nula sólo se compone de elementos nulos, es decir, es una matriz nula. Ésta es la única matriz que al mismo tiempo es simétrica y antisimétrica, al igual que la forma nula.

Ya se ha notado que existe una relación estrecha entre las formas bilineales simétricas y las cuadráticas. Esta relación se observa claramente en el nivel matricial. Para una forma bilineal $\varphi(x, y)$ es válida la correlación matricial (91.5). Para la forma cuadrática correspondiente tenemos

$$\varphi(x, x) = x_e' G_e x_e. \qquad (91.8)$$

Siendo fijada la base $e_1, e_2, \ldots, e_n$, cada anotación del tipo (91.8) define, para cualquier matriz G_e, cierta forma cuadrática. La matriz G_e en (91.8) ya no se llama matriz de la forma bilineal, sino *matriz de la forma cuadrática*.

Si para las formas bilineales existe una correspondencia biunívoca entre las formas y las matrices de éstas, siendo fijada una base en K_n, en las circunstancias que se describen tal correspondencia ya no existe. Cada forma cuadrática puede ser definida por un conjunto de sus matrices. Dicho conjunto contiene una sola matriz simétrica y la diferencia entre cualesquiera dos matrices del conjunto dado es una matriz antisimétrica.

De este modo, cualquier forma cuadrática arbitraria puede ser definida siempre por una matriz simétrica. Al pasar a otra base, las matrices de la forma cuadrática varían de conformidad con (91.6). Por esta razón concluimos otra vez que los problemas de investigación de las formas bilineales simétricas y de las cuadráticas están estrechamente entrelazados. Para las formas cuadráticas hermitianas ya no es así, puesto que entre éstas y las formas bilineales hermitianas existe una correspondencia biunívoca y la misma correspondencia se mantiene entre sus matrices.

Por analogía con las formas bilineales, se llamará *rango* de una forma cuadrática el rango de su matriz en cualquier base. Si la matriz de una forma cuadrática es regular, la forma cuadrática se denominará también *regular*.

El estudio de las formas bilineales significa, en esencia, el estudio de sus matrices en diferentes bases o, lo que es igual, el estudio de la clase de matrices congruentes. Por ello, las investigaciones que siguen quedarán relacionadas con el examen de las clases de las matrices congruentes y congruentes según Hermite.

Para las clases de este género pueden indicarse ahora mismo toda una serie de propiedades que provienen de los resultados obtenidos anteriormente. Así por ejemplo, una matriz congruente de la simétrica (antisimétrica) es necesariamente simétrica (antisimétrica). En particular, una matriz congruente de la matriz diagonal será simétrica. De aquí concluimos que una matriz simétrica no nula nunca será congruente de la antisimétrica, aunque pueda ser equivalente a ella. Una matriz antisimétrica no nula nunca puede ser congruente de la diagonal. Una matriz, congruente según Hermite de una matriz hermitiana (antihermitiana), es obligatoriamente

hermítica (antihermítica). Entre las matrices diagonales, como matriz hermítica (antihermítica) puede intervenir sólo la que tiene elementos reales (imaginarios puros).

En concordancia con las descomposiciones (90.4), (90.7) de las formas bilineales y bilineales según Hermite, obtenemos unas descomposiciones de una matriz arbitraria en la suma de las matrices simétrica y antisimétrica, así como también de las matrices hermítiana y antihermítiana. Estas descomposiciones pueden ser escritas en la forma explícita:

$$A = \frac{1}{2}(A + A') + \frac{1}{2}(A - A'),$$

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*).$$

Si A es una matriz de la forma bilineal, los primeros sumandos de los segundos miembros son matrices de las partes simétricas de la forma bilineal y los segundos, las matrices de las partes antisimétricas de la misma forma.

En adelante haremos extender frecuentemente, sin explicaciones adicionales, la terminología introducida para las formas bilineales y cuadráticas a las matrices. Por ejemplo, llamamos una matriz *definida positiva*, entendiendo por ello que es una matriz de una forma definida positiva, etc.

Uno de los problemas más importantes relacionados con la forma bilineal es la determinación de la expresión más simple a la que puede ser reducida la matriz de la forma dada cuando varía la base y la búsqueda de la base correspondiente. Este problema lleva el nombre de *transformación* de la forma bilineal o de *reducción* de la forma bilineal a una expresión más simple.

En la interpretación matricial el problema de transformación puede enunciarse de la manera siguiente:

Dada la matriz A, hállese tal matriz regular P que la matriz

$$C = P'AP, \tag{91.9}$$

congruente de A, tenga la forma más simple.

Esto nos da, de hecho, una descomposición de la matriz en factores, puesto que de (91.9) se deduce que

$$A = (P^{-1})' \, CP^{-1}.$$

Por supuesto, para las formas bilineales hermitianas, en lugar de (91.9) consideraremos las transformaciones

$$C = P'A\overline{P}. \tag{91.10}$$

Desde el punto de vista de los cálculos es importante que la matriz P en (91.9), (91.10) sea no muy compleja. Esto se debe a que al buscar las nuevas coordenadas de los vectores en términos de las antiguas, conforme a (63.3), nos vemos obligados a resolver un siste-

mo de ecuaciones algebraicas lineales con la matriz P y es menester que la resolución se realice lo suficientemente rápido. En algunos casos, en lugar de la matriz P resulta más cómodo buscar la matriz P^{-1}.

Además de las formas consideradas de escribir las formas bilineales y cuadráticas, se usan también algunas otras. A veces las definiremos explícitamente:

$$\Phi = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji} x_i y_j,$$

$$F = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji} x_i x_j. \qquad (91.11)$$

Estas anotaciones pueden ser simplificadas. Sea, por ejemplo, real un espacio, entonces reales serán también tanto la propia forma bilineal como la matriz A compuesta por los coeficientes a_{ji}. Introduzcamos el espacio $\mathbf{R}_n$ cuyos elementos son los vectores columna

$$x = (x_1, x_2, \ldots, x_n)', \quad y = (y_1, y_2, \ldots, y_n)'$$

y supongamos que el producto escalar se ha introducido como una suma de productos de las coordenadas tomadas dos a dos. Ahora podemos escribir

$$\Phi = (Ax, y), \quad F = (Ax, x). \qquad (91.12)$$

Para las formas bilineales hermitianas, anotadas como

$$\Phi = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji} x_i \overline{y}_j, \qquad F = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji} x_i \overline{x}_j,$$

se verifica también (91.12), si, desde luego, el producto escalar se introduce como una suma de productos de las coordenadas del primer vector por las coordenadas conjugadas complejas del segundo vector.

Ejercicios.

1. Demuéstrese que el determinante de una matriz hermitiana es un número real.
2. ¿Qué número es el determinante de una matriz antihermitiana?
3. Demuéstrese que el rango de una matriz antisimétrica es un número par.
4. Las formas bilineales $\varphi(x, y)$ y $\overline{\varphi(y, x)}$ son, en general, diferentes. ¿Qué puede decirse sobre sus matrices?
5. Demuéstrese que el rango de la suma de unas formas bilineales no es superior a la suma de los rangos de los sumandos.
6. Demuéstrese que se puede representar toda forma bilineal de rango r como suma de r formas bilineales de rango 1.
7. Demuéstrese que se puede representar toda forma bilineal $\varphi(x, y)$ de rango 1 como

$$\varphi(x, y) = \varphi(x, a)\,\varphi(b, y)$$

para ciertos vectores a, b. ¿Será única tal representación?

§ 92. Reducción a una forma canónica

Antes de empezar a investigar las diferentes esferas del empleo de las formas bilineales y cuadráticas, consideraremos un método general de la transformación congruente o congruente según Hermite, de las matrices a una forma sencilla.

Sea dada una matriz cuadrada A de orden n y se necesita hallar tal matriz regular P que la matriz $C = P'AP$ tenga una forma suficientemente sencilla. Realizándose una transformación congruente según Hermite, una forma sencilla la debe poseer la matriz $C = P'A\overline{P}$. Expondremos ahora un método general de la transformación que será útil para todas las matrices A. La diferencia entre las transformaciones congruente y congruente según Hermite será insignificante. Por ello, para concretar, convengamos en considerar que se realiza la transformación congruente de una matriz.

El método consiste en construcción de una sucesión de matrices $A_0 = A, A_1, A_2, \ldots, A_s$ en la que cada matriz consecutiva es congruente de la anterior, es decir,

$$A_{k+1} = P'_{k+1}A_k P_{k+1}$$

para cierta matriz P_{k+1}. Puesto que la relación de congruencia es transitiva, la última matriz A_s será congruente de la matriz inicial A. El principio de construcción de la sucesión de matrices A_k está fundado en que para todo k se obtengan en la matriz A_{k+1} más elementos nulos que en la matriz A_k. Más aún, cada vez, al calcular la matriz P_{k+1} según la A_k, exigiremos que en la matriz A_{k+1} no sólo aparezcan nuevos elementos nulos, sino que se guarden todos los elementos nulos obtenidos en todas las etapas anteriores.

La transformación de una matriz A_k en la A_{k+1} se llamará paso principal del método. Cada paso principal puede consistir en varios pasos auxiliares. Todos ellos se reducirán a la ejecución de las operaciones elementales: la permutación de las columnas (filas) de una matriz, la adición a una columna (fila) de otra columna (fila) multiplicada por un número, la multiplicación de una columna (fila) por un número. Describiremos los pasos auxiliares en términos de las transformaciones de la matriz A en otra matriz, congruente de ella, $C = P'AP$, omitiendo, para simplificar, el índice k.

A. En la matriz A el elemento $a_{11} \neq 0$. Existe una matriz regular P tal que para los elementos de la primera columna de la matriz $C = P'AP$ se verifican las correlaciones

$$c_{j1} = \begin{cases} a_{11}, & j = 1, \\ 0, & j \neq 1. \end{cases} \qquad (92.1)$$

La matriz P difiere de la matriz unidad sólo en su primera fila, con la particularidad de que

$$p_{1j} = \begin{cases} 1, & j = 1, \\ -\dfrac{a_{1j}}{a_{11}}, & j \neq 1. \end{cases} \qquad (92.2)$$

La multiplicación a la izquierda de la matriz A por P' no altera la primera fila de la matriz A y convierte en cero todos los elementos de la primera columna de la matriz $P'A$ dispuestos fuera de la diagonal. La multiplicación a la derecha de la matriz $P'A$ por P no cambia la primera columna de la matriz $P'A$.

Hemos de señalar una circunstancia más. Llamaremos *principales* a todos los menores de la matriz dispuestos en la esquina izquierda superior. Como la matriz P es triangular derecha y todos los elementos diagonales de ella son iguales a la unidad, de todos los menores dispuestos en las primeras r columnas será distinto de cero sólo el menor principal; es igual a la unidad. Por esta razón en las matrices A y C coincidirán todos los menores principales. En efecto, haciendo uso de la fórmula Binet—Cauchy, obtenemos

$$C\begin{pmatrix} 1 & 2 \dots r \\ 1 & 2 \dots r \end{pmatrix} = \sum_{1 \le k_1 < k_2 < \dots < k_r \le n} P'\begin{pmatrix} 1 & 2 \dots r \\ k_1 & k_2 \dots k_r \end{pmatrix} \times$$

$$\times AP\begin{pmatrix} k_1 & k_2 \dots k_r \\ 1 & 2 \dots r \end{pmatrix} = AP\begin{pmatrix} 1 & 2 \dots r \\ 1 & 2 \dots r \end{pmatrix} =$$

$$= \sum_{1 \le k_1 < k_2 < \dots < k_r \le n} A\begin{pmatrix} 1 & 2 \dots r \\ k_1 & k_2 \dots k_r \end{pmatrix} \cdot P\begin{pmatrix} k_1 & k_2 \dots k_r \\ 1 & 2 \dots r \end{pmatrix} =$$

$$= A\begin{pmatrix} 1 & 2 \dots r \\ 1 & 2 \dots r \end{pmatrix}.$$

Esta observación la emplearemos más adelante.

B. En la matriz A el elemento a_{11} es igual a cero, pero cierto elemento a_{jj} es distinto de 0, $j > 1$. Existe una matriz regular P tal que para la matriz $C = P'AP$ el elemento $c_{11} = a_{jj}$ es diferente de cero. La matriz P se diferencia de la matriz unidad sólo en cuatro elementos dispuestos en la intersección de las filas y las columnas con números $1, j$. En estas posiciones la matriz P tiene por expresión $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. La multiplicación a la derecha de la matriz A por P permuta en la matriz A las columnas con los números $1, j$. La multiplicación de la matriz AP a la izquierda por la matriz P' permuta en la matriz AP las filas con los números $1, j$.

C. En la matriz A todos los elementos diagonales son nulos, pero hay tales índices j, l, donde $j < l$, que $a_{jl} + a_{lj} \neq 0$. Existe una

matriz regular P tal que para la matriz $C = P'AP$ el elemento $c_{jj} = a_{lj} + a_{jl}$ es distinto de 0. La matriz P se diferencia de la matriz unidad en el elemento $p_{lj} = 1$. La multiplicación a la derecha de la matriz A por P agrega a la j-ésima columna de la matriz A su l-ésima columna. La multiplicación a la izquierda de la matriz AP por P' agrega a la j-ésima fila de la matriz AP su l-ésima fila.

D. La matriz A es antisimétrica no nula, el elemento a_{12} es igual a 0, pero cierto elemento a_{jl} es distinto de 0, donde $j < l$. Existe una matriz regular P tal que en la matriz antisimétrica $C = P'AP$ el elemento $c_{12} = a_{jl}$ es diferente de 0. La matriz P viene representada como el producto $P = P_1 \cdot P_2$. Las matrices P_1, P_2 se diferencian de las matrices unidad sólo en cuatro elementos dispuestos en la intersección de las filas y las columnas con los números respectivos 1, j y 2, l. En estas posiciones las matrices P_1 y P_2 tienen por expresión $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Como ya se ha dicho, la multiplicación a la derecha por estas matrices conlleva la permutación de las columnas, la multiplicación a la izquierda, la permutación de las filas.

E. La matriz del menor principal de tercer orden de la matriz A tiene por expresión

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{pmatrix}, \qquad (92.3)$$

donde los elementos a_{11}, a_{23} y a_{32} son distintos de cero. Existe una matriz regular P tal que en la matriz $C = P'AP$ serán diferentes de cero los primeros tres menores principales. La matriz P se diferencia de la matriz unidad en un elemento p_{31} el cual puede ser cualquier número, salvo 0, $-a_{11}a_{13}^{-1}$ y $-a_{13}a_{31}^{-1}$. La multiplicación a la derecha de la matriz A por P agrega a la primera columna de la matriz A su tercera columna multiplicada por p_{31}. La multiplicación a la izquierda de la matriz AP por P' agrega a la primera fila de la matriz AP su tercera fila multiplicada por p_{31}.

F. La matriz A es antisimétrica, el elemento a_{12} es distinto de 0. Existe una matriz regular P tal que para los elementos de las primeras dos columnas de la matriz $C = P'AP$ se verifican las correlaciones

$$c_{j1} = \begin{cases} -a_{12}, & j = 2, \\ 0, & j \neq 2, \end{cases} \quad c_{j2} = \begin{cases} a_{12}, & j = 1, \\ 0, & j \neq 1. \end{cases}$$

Puesto que en una transformación congruente una matriz antisimétrica se transforma en otra, también antisimétrica, las correlaciones análogas tendrán lugar también para las primeras dos filas

de la matriz C. La matriz P se representa como un producto $P = P_1 \cdot P_2$. La matriz P_1 se diferencia de la matriz unidad sólo en la segunda fila, siendo, además,

$$p^{(1)}_{2j} = \begin{cases} 0, & j = 1, \\ 1, & j = 2, \\ \frac{a_{1j}}{a_{12}}, & j > 2. \end{cases}$$

La matriz P_2 se diferencia de la matriz unidad sólo en la primera fila, verificándose en este caso.

$$p^{(2)}_{1j} = \begin{cases} 1, & j = 1 \\ 0, & j = 2, \\ -\frac{a_{2j}}{a_{11}}, & j > 2. \end{cases}$$

La multiplicación a la izquierda de la matriz A por P_1' no altera las primeras dos filas y la segunda columna de la matriz A, convirtiendo en cero todos los elementos en la primera columna de la matriz $P_1'A$, a excepción de los dos primeros. La multiplicación a la izquierda de la matriz $P_1'A$ por P_2' no altera las primeras dos filas y la primera columna de la matriz $P_1'A$, convirtiendo en cero todos los elementos de la segunda columna de la matriz $P'A$, a excepción de los dos primeros. La multiplicación a la derecha de la matriz $P'A$ por P no cambia las primeras dos columnas de la matriz $P'A$.

6. Supongamos que la matriz A tiene, realizada cierta partición en células, la siguiente estructura

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right), \qquad (92,6)$$

donde A_{11}, A_{11} son células cuadradas. Si P_{22} es una matriz regular cuyo orden es igual al de A_{22}, entonces la matriz

$$C = \left(\begin{array}{c|c} A_{11} & A_{12}P_{22} \\ \hline 0 & P'_{22}A_{22}P_{22} \end{array}\right)$$

es congruente de la matriz A. En este caso $C = P'AP$, donde

$$P = \left(\begin{array}{c|c} E & 0 \\ \hline 0 & P_{22} \end{array}\right).$$

La comprobación directa de todas las afirmaciones enunciadas al describir los pasos auxiliares no representa alguna dificultad singular y por esto proponemos que el lector mismo se convenza de su veracidad lo que puede hacerse en calidad de ejercicios.

El método, en total, se realiza del modo siguiente. En el primer paso principal la matriz A se reduce a la forma (92.4), donde A_{11} es una matriz regular de orden uno o dos. Si la matriz A_k, $k \geqslant 1$, tiene la forma (92.4), entonces, al realizarse el segundo paso principal, la matriz en la esquina inferior derecha se reduce también a la forma (92.4) y se lleva a cabo la transformación congruente general de conformidad con el paso G. La matriz A_{k+1} puede, nuevamente, representarse en la forma (92.4), pero la célula en la esquina izquierda superior para A_{k+1} no sólo será regular, sino tendrá el orden mayor en comparación con el de la matriz A_k. El proceso se repite hasta que, realizado algún paso, en la representación (92.4) de la matriz A_k aparezca una célula nula en la esquina inferior derecha o bien el orden de la célula en la esquina superior izquierda se haga igual a n. En este caso, la matriz de la transformación resultante será igual al producto de izquierda a derecha de las matrices de transformaciones de todos los pasos.

La forma de la matriz A_k depende de si es o no la matriz A antisimétrica. De esto mismo depende también de qué modo los pasos auxiliares forman parte de los pasos principales del método.

Cualquiera que sea el paso principal, su finalidad consiste en obtener una porción seguida de ceros en la matriz a transformar. Si la matriz inicial no es antisimétrica, los ceros se obtienen siempre con ayuda del paso auxiliar A, mientras que los pasos B — C sólo se necesitan para preparar A. En cambio, si la matriz inicial es antisimétrica, los ceros se obtienen con ayuda del paso F, mientras que D será paso auxiliar. Describamos también el paso principal del método en términos de la transformación de la matriz A y empecemos con la matriz no antisimétrica A.

En el primer paso principal, la matriz que se transforma no es antisimétrica. Si el elemento $a_{11} \neq 0$ y todos los elementos extradiagonales de la primera columna son nulos, entonces nada varía y consideramos terminado el paso principal. A título de matriz de la transformación P tomamos, en este caso, la matriz unidad. En el caso general realizamos el primero de los pasos auxiliares A — C, que puede ser llevado a cabo. Si tal paso resulta ser B o C, después de éste se cumple obligatoriamente el paso A o bien ambos pasos B, A. A título de matriz de la transformación P tomamos el producto de izquierda a derecha de todas las matrices de transformaciones de los pasos auxiliares realmente realizados. Después de cumplir el primer paso principal, en la matriz transformada A_1 todos los elementos extradiagonales de la primera columna serán nulos, es decir, la matriz A_1 será de estructura celular del tipo (92.4).

La diferencia de todos los pasos restantes con respecto al primero está relacionada con el hecho de que la matriz a transformar puede resultar antisimétrica. Si ésta no es antisimétrica, el paso principal siguiente no difiere en nada del paso primero. En cambio, si la

matriz que se transforma es antisimétrica, entonces cualquiera que sea su transformación congruente, queda antisimétrica y, sirviéndose sólo de esta matriz, no se puede obtener un elemento no nulo en la esquina superior izquierda. La salida de esta situación está basada en la necesidad de transformar la célula diagonal inferior ampliada.

Hasta que se encuentre una matriz antisimétrica, la célula en la esquina superior izquierda de la representación (92.4) para las matrices A_k será triangular derecha con elementos diagonales no nulos. Si los elementos en las posiciones (1, 2) y (2, 1) de la matriz antisimétrica en la esquina inferior derecha son distintos de cero, entonces para la matriz A_r, la siguiente en la transformación, sustituyamos la representación (92.4), disminuyendo en uno el orden de la célula en la esquina superior izquierda. Ahora la matriz de tercer orden en la esquina superior izquierda de la nueva célula diagonal inferior tendrá la forma (92.3) y se puede realizar el paso auxiliar **E**. Después de esto podemos realizar tres veces consecutivas el paso **A**. Efectivamente, según lo observado, la ejecución del paso **A** no cambia los menores principales de la matriz. Por consiguiente, en el caso dado, realizado el paso **A**, la nueva matriz en la esquina inferior derecha tendrá distintos de cero los dos primeros menores principales. Por ello, podemos, a ciencia cierta, hacer un paso **A** más. Los razonamientos análogos demuestran que el paso **A** puede realizarse también por tercera vez. Al retroceder un paso "atrás" hemos obtenido la posibilidad de avanzar tres pasos "adelante". Cuando sea necesario, antes de realizar el paso **E**, se lleva a cabo el paso **D**.

De suerte, si la matriz A no es antisimétrica, el método que acabamos de exponer permite construir una matriz regular P tal que la matriz $P'AP$, congruente de A, tendrá la estructura siguiente:

$$P'AP=\left(\begin{array}{c|c} M & N \\ \hline 0 & 0\end{array}\right). \qquad (92.5)$$

Aquí M es una matriz triangular derecha con los elementos diagonales no nulos, el orden de la matriz M es igual al rango de la matriz A.

Si A es una matriz antisimétrica, todos los pasos principales del método, incluido el primero, se realizan siguiendo un mismo esquema. Supongamos que ya se ha obtenido la matriz A_k del tipo (92.4), con la particularidad de que en la esquina superior izquierda se dispone una matriz diagonal celular regular de células antisimétricas de segundo orden. Dado que, al realizar una transformación congruente, una matriz antisimétrica se transforma en otra antisimétrica, entonces la célula A_{12} en (92.4) será nula. Primero se logra que en las posiciones (1, 2) y (2, 1) de la matriz antisimétrica la esquina inferior derecha sea ocupada por elementos no nulos. Es posible que para esto resulte necesario realizar el paso auxiliar **D**. Luego realizamos el paso **F**, lo que adjunta a la diagonal una célula

más que es antisimétrica regular de segundo orden. A continuación, pasamos al siguiente paso principal. En este caso también el proceso continúa hasta que, realizado cierto paso, en la representación (92.4) de la matriz A_r aparezca una célula nula en la esquina inferior derecha, o bien el orden de la célula en la esquina superior izquierda se haga igual a n.

Así pues, si la matriz A es antisimétrica, el método permite construir una matriz regular P, para la cual la matriz $P'AP$ será de la estructura siguiente:

$$P'AP=\left(\begin{array}{c|c} M & 0 \\ \hline 0 & 0 \end{array}\right) \qquad (92.6)$$

Aquí M es una matriz diagonal celular con células antisimétricas regulares de segundo orden. El orden de la matriz M es igual al rango de la matriz A.

En la transformación congruente hermitiana el esquema general del método queda inalterable. Sin embargo, el propio proceso resulta ser más fácil, comparando con la transformación congruente ordinaria, y el paso auxiliar C se sustituye por el siguiente.

C'. En la matriz A todos los elementos diagonales son nulos, pero hay tales índices j, l, donde $j < l$, que entre los elementos a_{jl}, a_{lj} existe aunque sea uno diferente de cero. Existe una matriz regular P tal que para la matriz $C = P'A\bar{P}$ uno de los elementos diagonales c_{jj}, c_{ll} es distinto de cero. A saber, $c_{jj} = a_{lj} + a_{jl}$, $c_{ll} = i(a_{lj} - a_{jl})$. La matriz P se diferencia de la matriz unidad en dos elementos $p_{lj} = 1$, $p_{jl} = i$. La multiplicación a la derecha de la matriz A por $\bar{P}$ agrega a la j-ésima columna de la matriz A su l-ésima columna y a la l-ésima columna, su j-ésima columna multiplicada por $-i$. La multiplicación a la izquierda de la matriz $A\bar{P}$ por P' adjunta a la j-ésima fila de la matriz $A\bar{P}$ su l-ésima fila y a la l-ésima fila, su j-ésima fila multiplicada por i.

Ahora no hay necesidad en los pasos D — F del método general, puesto que nunca sobrepasaremos el paso C'. Además, las fórmulas (92.2) quedan intactas.

De esta modo, si A es una matriz no nula, el método hace posible construir tal matriz regular P que la matriz $P'A\bar{P}$, congruente de A según Hermite, será de la siguiente estructura:

$$P'A\bar{P}=\left(\begin{array}{c|c} M & N \\ \hline 0 & 0 \end{array}\right). \qquad (92.7)$$

Aquí, M es una matriz triangular derecha con elementos diagonales no nulos. El orden de la matriz M es igual al rango de la matriz A.

Los tipos de las matrices (92.5) — (92.7) se denominan formas *canónicas* para las operaciones de la transformación congruente. Se

llama *canónica* también cualquier base en la que la matriz inicial tiene la forma indicada. Las propias matrices del tipo (92.5), (92.7) llevan el nombre de *trapezoidales derechas*. De modo análogo se definen las matrices *trapezoidales izquierdas*.

Demos a conocer algunas deducciones interesantes que provienen de las formas canónicas de matrices. Ya hemos dicho que en una transformación congruente se conserva el carácter simétrico y antisimétrico de la matriz. Si una de estas propiedades la poseía la matriz inicial, debe quedarse válida para la forma canónica. Por esto, en adición a lo dicho podemos concluir que

Una matriz simétrica es congruente de la matriz diagonal.

Una matriz hermitiana es congruente según Hermite de la matriz diagonal real.

Una matriz antihermitiana es congruente según Hermite de la matriz diagonal imaginaria pura.

En todos estos casos la reducción a una forma canónica se efectúa con una facilidad singular, puesto que no puede surgir la necesidad de llevar a cabo aunque sea uno solo de los pasos auxiliares D — F.

Sobre las matrices de la forma canónica del tipo (92.5), (92.6) se puede realizar una transformación congruente con la matriz diagonal más y conseguir que los elementos no nulos que determinan la regularidad de la célula M sean iguales a $+1$ o bien a -1. Esta forma canónica de la matriz y la base que le corresponde se denominan *normales*. Está claro que la multiplicación a la derecha (a la izquierda) por una matriz diagonal conduce a la multiplicación de las columnas (filas) por los elementos diagonales de la matriz de la transformación. Describamos nuevamente esta transformación en términos del cumplimiento del paso auxiliar con la matriz A.

H. Una matriz real no antisimétrica A de rango r es de la forma canónica (92.5). Existe una matriz diagonal real P tal que los elementos diagonales no nulos c_{jj} de la matriz $C = P'AP$ son iguales a sign a_{jj}. Además

$$p_{jj}=\begin{cases}(a_{jj}\,\mathrm{sign}\,a_{jj})^{-1/2}, & j\leqslant r,\\ 1, & j>r.\end{cases}$$

Una matriz real (compleja) antisimétrica A de rango r es de la forma canónica (92.6). Existe una matriz diagonal real (compleja) P tal que los elementos no nulos dispuestos por arriba de la diagonal de la matriz $C = P'AP$ son iguales a $+1$, y los elementos no nulos dispuestos por debajo de la diagonal, iguales a -1. En este caso

$$p_{jj}=\begin{cases}1, & j \text{ es impar},\\ a_{j-1,\,j}^{-1}, & j \text{ es par}.\end{cases}$$

Una matriz compleja no antisimétrica A de rango r tiene la forma canónica (92.5). Existe una matriz diagonal compleja P tal que los elementos diagonales no nulos c_{jj} de la matriz $C = P'AP$ son iguales

a uno. En este caso

$$p_{jj} = \begin{cases} a_{jj}^{-1/2}, & j \leqslant r, \\ 1, & j > r. \end{cases}$$

La transformación congruente según Hermite con una matriz diagonal se efectúa raras veces, puesto que con su ayuda sólo pueden cambiarse los módulos de los elementos que determinan la regularidad de la célula M en (92.7), pero no se puede hacer reales los elementos diagonales complejos.

Ejercicios.

1. Demuéstrese que si la reducción a la forma canónica mediante la matriz P se efectúa según el método descrito más arriba, entonces $\det P = \pm 1$.

2. ¿Qué significa, desde el punto de vista de la forma canónica, la igualdad matricial

$$\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -i \end{pmatrix} = \begin{pmatrix} 1+i & 0 \\ 2 & 1+i \end{pmatrix} ? \qquad (92.8)$$

3. ¿A qué forma puede reducirse una matriz no simétrica con ayuda de una transformación congruente, si se excluye el paso auxiliar E?

4. ¿Qué forma tiene la matriz P de una transformación, si cada paso principal del método considerado más arriba consiste sólo en el paso auxiliar A?

5. Demuéstrese que toda matriz triangular derecha es congruente de la matriz triangular izquierda. ¿Cuál es la forma más simple de la matriz de una transformación?

6. Demuéstrese que toda matriz regular de orden impar es congruente de la matriz triangular derecha regular.

7. Sea G una matriz de una forma bilineal definida positiva. Demuéstrese que para sus elementos g_{ij} se verifican las correlaciones

$$g_{ii} > 0, \qquad (g_{ij} + g_{ji})^2 < 4 g_{ii} g_{jj},$$

cualesquiera que sean i, j.

8. Sea G una matriz de una forma bilineal definida negativa. Demuéstrese que para sus elementos g_{ij} se verifican las correlaciones

$$g_{ii} < 0, \qquad (g_{ij} + g_{ji})^2 < 4 g_{ii} g_{jj},$$

cualesquiera que sean i, j.

9. Demuéstrese que las matrices de todas las formas bilineales simétricas definidas positivas (negativas) son congruentes entre sí mismas.

10. Demuéstrese que para que G sea una matriz de la forma bilineal de signo variable, es suficiente que entre sus elementos diagonales haya elementos de signos distintos.

§ 93. Congruencia y descomposiciones matriciales

El método general de la transformación congruente de una matriz a la forma canónica no siempre permite decir de antemano, cuál será la matriz de la transformación de coordenadas al pasar a la base canónica. No obstante, con ciertas

restricciones adicionales impuestas en la matriz inicial, para dicha pregunta existe una respuesta bien determinada.

Supongamos que la matriz A tiene distintos de cero todos los menores principales, a excepción, quizás, del menor de orden superior, es decir, del determinante de la matriz A. Probemos que tal matriz siempre puede representarse en forma del producto

$$A = LDU, \tag{93.1}$$

donde L es una matriz triangular izquierda con elementos diagonales unidades, D es una matriz diagonal y U, una matriz triangular derecha con elementos diagonales unidades, es decir,

$$A = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & 0 \\ & & \ddots & \\ l_{n1} & l_{n2} & \dots & 1 \end{pmatrix} \begin{pmatrix} d_{11} & & & \\ & d_{22} & & 0 \\ & & \ddots & \\ 0 & & & d_{nn} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & \dots & u_{1n} \\ & 1 & \dots & u_{2n} \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}.$$

Igualando entre sí los elementos de la matriz A y del producto LDU obtenemos

$$a_{ij} = \sum_{p=1}^{\min(i,\, j)} l_{ip} d_{pp} u_{pj}. \tag{93.2}$$

Ahora, de (93.2) hallamos de manera sucesiva todos los elementos desconocidos de las matrices de la descomposición (93.1). A saber,

$$\begin{gathered} d_{11} = a_{11}, \\ u_{1j} = \frac{a_{1j}}{d_{11}}, \quad l_{j1} = \frac{a_{j1}}{d_{11}}, \quad j > 1, \\ d_{ii} = a_{ii} - \sum_{p=1}^{i-1} l_{ip} d_{pp} u_{pi}, \quad i > 1, \\ u_{ij} = \frac{a_{ij} - \sum_{p=1}^{i-1} l_{ip} d_{pp} u_{pj}}{d_{ii}}, \\ l_{ji} = \frac{a_{ji} - \sum_{p=1}^{i-1} l_{jp} d_{pp} u_{pi}}{d_{ii}}, \quad i > 1, \quad j > i. \end{gathered} \tag{93.3}$$

Apliquemos a la correlación (93.1) la fórmula de Binet—Cauchy. Recordemos que entre los menores de la matriz triangular izquierda L, dispuestos en las primeras r filas, sólo el menor principal es diferente de cero: es igual a la unidad. Los razonamientos análogos tienen lugar también para la matriz U, al cambiar entre sí, por su-

puesto, las filas y las columnas. Por ello

$$A\begin{pmatrix}1 & 2 \dots r\\ 1 & 2 \dots r\end{pmatrix} = \sum_{1 \le k_1 < k_2 < \dots < k_r \le n} L\begin{pmatrix}1 & 2 & \dots & r\\ k_1 & k_2 & \dots & k_r\end{pmatrix} \times$$

$$\times DU\begin{pmatrix}k_1 & k_2 \dots k_r\\ 1 & 2 \dots r\end{pmatrix} = DU\begin{pmatrix}1 & 2 \dots r\\ 1 & 2 \dots r\end{pmatrix} = d_{11}d_{22}\dots d_{rr}.$$

De aquí concluimos que

$$d_{11} = a_{11}, \qquad d_{ii} = \frac{A\begin{pmatrix}1 & 2 \dots i\\ 1 & 2 \dots i\end{pmatrix}}{A\begin{pmatrix}1 & 2 \dots i-1\\ 1 & 2 \dots i-1\end{pmatrix}}. \tag{93.4}$$

Por hipótesis, los menores principales de la matriz A son distintos de cero. Por consiguiente, serán distintos de cero todos los elementos diagonales d_{ii} en (93.4), salvo, quizás, el último elemento.

Trataremos con frecuencia las descomposiciones (93.1) para las matrices simétricas y hermitianas. Si esta vez también la matriz A tiene sus menores principales distintos de cero, a excepción, quizás, del último, entonces la matriz simétrica siempre puede ser representada en forma del producto

$$A = S'DS, \tag{93.5}$$

y la matriz hermitiana, en forma del producto

$$A = S'D\bar{S}. \tag{93.6}$$

Aquí S es una matriz triangular derecha cuyos elementos diagonales son iguales a uno, D es una matriz diagonal, es decir,

$$S = \begin{pmatrix}1 & s_{12} \dots s_{1n}\\ & 1 \dots s_{2n}\\ & \ddots \\ 0 & & 1\end{pmatrix} \qquad D = \begin{pmatrix}d_{11} & & 0\\ & d_{22} & \\ & & \ddots \\ 0 & & d_{nn}\end{pmatrix}$$

En concordancia completa con (93.3) tendremos ahora

$$d_{11} = a_{11}, \quad s_{1j} = \frac{a_{1j}}{d_{11}}, \quad j > 1,$$

$$d_{ii} = a_{ii} - \sum_{p=1}^{i-1} d_{pp}s_{pi}^2, \quad i > 1, \tag{93.7}$$

$$s_{ij} = \frac{a_{ij} - \sum_{p=1}^{i-1} d_{pp}s_{pi}s_{pj}}{d_{ii}}, \quad j > i.$$

para la descomposición (93.5) y

$$d_{11}=a_{11}, \quad s_{1j}=\frac{\bar a_{1j}}{d_{11}}, \quad j>1,$$

$$d_{ii}=a_{ii}-\sum_{p=1}^{i-1} d_{pp}\,|s_{pi}|^2, \quad i>1,$$

$$s_{ij}=\frac{a_{ij}-\sum\limits_{p=1}^{i-1} d_{pp}s_{pi}\bar s_{pj}}{d_{ii}}, \quad j>i,$$

para la descomposición (93.6). En este caso las fórmulas (93.4) siguen siendo válidas.

Las descomposiciones (93.4), (93.5), (93.6) son de amplio uso en la resolución de los más diversos problemas del álgebra lineal. En lo que se refiere a las transformaciones congruentes de una matriz, los datos de la descomposición conducen a las siguientes correlaciones:

$$(L^{-1'})'AL^{-1'}=DUL^{-1'}, \qquad (L^{-1'})'A\overline{L^{-1'}}=DU\overline{L^{-1'}},$$

$$S^{-1'}AS^{-1}=D, \qquad S^{-1'}A\bar S^{-1}=D.$$

Las matrices $DUL^{-1'}$ y $DU\overline{L^{-1'}}$ son triangulares derechas, las D son diagonales, con la particularidad de que en sus diagonales principales solamente el último elemento puede ser nulo. De modo que otra vez hemos obtenido los tipos ya conocidos de las matrices en una transformación congruente. Sin embargo, ahora se puede afirmar que las matrices de la transformación de coordenadas, al pasar a la base canónica, serán triangulares derechas, puesto que lo son las matrices $L^{-1'}$, S^{-1}. Las descomposiciones examinadas proporcionan las matrices L', S de las transformaciones de coordenadas, al pasar de la base canónica a la inicial, las cuales asimismo serán triangulares derechas.

En el caso de una matriz simétrica el proceso descrito de descomposición está estrechamente relacionado con el así llamado *algoritmo de Jacobi* de la transformación de una forma cuadrática a la forma canónica. La diferencia sólo consiste en que el algoritmo de Jacobi tiene determinada la matriz S^{-1}, en lugar de la matriz S. Ha de señalarse que la matriz S se halla de un modo más simple que S^{-1}.

Las transformaciones congruentes con una matriz triangular derecha son las más sencillas, no obstante lo suficientemente generales todavía para que puedan ser aplicadas a una clase amplia de matrices. Por esta razón causa un interés determinado la descripción de aquella clase de matrices que pueden reducirse a la forma canónica con ayuda de una transformación con la matriz triangular derecha.

LEMA 93.1. *Si una matriz rectangular A está representada en la forma celular*

$$A=\left(\frac{B \mid Q}{R \mid T}\right), \qquad (93.8)$$

donde B es una matriz cuadrada regular de orden r, entonces el rango de la matriz A es igual a r, si, y sólo si,

$$T = RB^{-1}Q. \qquad (93.9)$$

DEMOSTRACIÓN. Multipliquemos la matriz A a la izquierda por una matriz celular regular

$$P=\left(\frac{E \mid 0}{-RB^{-1} \mid E}\right),$$

donde las células correspondientes tienen las mismas dimensiones que en (93.8). Entonces

$$PA=\left(\frac{B \mid Q}{0 \mid T-RB^{-1}Q}\right).$$

Las matrices A y PA son de un mismo rango el cual será igual a r cuando, y sólo cuando, $T - RB^{-1}Q = 0$.

Ahora podemos describir la clase buscada de matrices. Resulta estrechamente relacionada con las matrices del tipo (93.8), (93.9).

TEOREMA 93.1. *Para que una matriz no antisimétrica pueda ser reducida a la forma canónica mediante la transformación congruente con una matriz triangular derecha, es necesario y suficiente que el número de los primeros menores principales no nulos de la matriz A sea igual a su rango.*

DEMOSTRACIÓN. NECESIDAD. Supongamos que una matriz no antisimétrica A se reduce, con ayuda de la matriz triangular derecha P, a la forma canónica (92.5). Es evidente que el número de los primeros menores principales no nulos de la matriz A no puede ser superior al orden de la célula M. Aplicando la fórmula de Binet—Cauchy y teniendo presente que en las primeras columnas de la matriz P el menor no nulo está ausente, salvo el principal, obtenemos

$$A\begin{pmatrix}1 & 2 \dots s\\ 1 & 2 \dots s\end{pmatrix}\left\{P\begin{pmatrix}1 & 2 \dots s\\ 1 & 2 \dots s\end{pmatrix}\right\}^2 = M\begin{pmatrix}1 & 2 \dots s\\ 1 & 2 \dots s\end{pmatrix}$$

para todo s no superior al orden de la matriz M. Como los menores principales de la matriz M y P son distintos de cero, el número de los primeros menores principales no nulos de la matriz A es igual a su rango.

SUFICIENCIA. Supongamos que el número de los primeros menores principales no nulos de la matriz A y el rango de ésta son iguales a r. Representemos la matriz A en la forma celular (93.8), donde el or-

den de la célula B es igual a r. Puesto que todos los menores principales de la matriz B son distintos de cero, entonces, de acuerdo con lo dicho más arriba, se la puede representar en la forma $B = LDU$, análogamente a (93.1). Construyamos una matriz celular

$$P = \left(\begin{array}{c|c} L^{-1'} & -B^{-1'}R' \\ \hline 0 & E \end{array}\right).$$

La comprobación directa muestra que

$$P'AP = \left(\begin{array}{c|c} DUL^{-1'} & L^{-1}(-RB^{-1'}R' + Q) \\ \hline 0 & 0 \end{array}\right).$$

La matriz $DUL^{-1'}$ es triangular derecha regular, la matriz P es triangular derecha regular y, por lo tanto, la matriz A se reduce a la forma canónica de un modo adecuado.

Para la transformación congruente de una matriz antisimétrica y la transformación congruente hermitiana de una matriz arbitraria las afirmaciones correspondientes se demuestran análogamente y aquí nos limitamos sólo a enunciarlas.

TEOREMA 93.2. *Para que una matriz antisimétrica A de rango r pueda ser reducida a la forma canónica mediante la transformación congruente con una matriz triangular derecha, es necesario y suficiente que el número de los primeros menores principales no nulos del orden par de la matriz A sea igual a $r/2$.*

TEOREMA 93.3. *Para que una matriz A pueda ser reducida a la forma canónica mediante la transformación congruente hermitiana con una matriz triangular derecha, es necesario y suficiente que el número de los primeros menores principales no nulos de la matriz A sea igual al rango de ésta.*

Las transformaciones de una matriz, tanto congruentes como congruentes según Hermite, no son, en el caso general, transformaciones de semejanza. No obstante, si para cierta clase de matrices P se verifica uno de los grupos de las correlaciones

$$PP' = P'P = E, \qquad PP^* = P^*P = E, \tag{93.10}$$

entonces, en este caso la transformación de congruencia se convierte en la de semejanza y para realizar las investigaciones se pueden utilizar los resultados obtenidos anteriormente referentes a la semejanza de matrices. Como ya sabemos, las matrices ortogonales reales satisfacen el primer grupo de las correlaciones en (93.10), las matrices unitarias complejas, el segundo grupo de correlaciones. Por eso, al recordar los resultados de los §§ 76—81, referentes a las semejanzas unitaria y ortogonal, concluimos que son justas las afirmaciones siguientes.

Toda matriz real, simétrica o antisimétrica, se reduce a la forma canónica mediante la transformación congruente con una matriz ortogonal.

Toda matriz compleja se reduce a la forma canónica por medio de la transformación congruente hermítiana con una matriz unitaria.

Estas afirmaciones son, en lo principal, de interés teórico, dado que en la práctica resulta muy difícil hallar matrices unitarias y ortogonales de la transformación, sobre todo cuando $n \geqslant 5$.

Ejercicios.

1. Demuéstrese que si las descomposiciones (93.1), (93.5), (93.6) existen, son únicas.

2. Demuéstrese que si todos los menores (a excepción, quizás, del menor de orden superior) de la matriz A, dispuestos en la esquina inferior derecha, son distintos de cero, entonces existe, y, además, sólo una descomposición $A = LDU$, donde L es una matriz triangular derecha, U, una matriz triangular izquierda con elementos diagonales iguales a uno, D, una matriz diagonal.

3. Demuéstrese que para los elementos d_{ii} de la matriz D que figura en el ejercicio 2, son válidas las correlaciones

$$d_{nn} = a_{nn}, \qquad d_{ii} = \frac{A\begin{pmatrix} i, & i+1, & \ldots, & n \\ i, & i+1, & \ldots, & n \end{pmatrix}}{A\begin{pmatrix} i+1, & i+2, & \ldots, & n \\ i+1, & i+2, & \ldots, & n \end{pmatrix}}, \quad i < n.$$

4. ¿En qué factores triangulares puede descomponerse una matriz, si son distintos de cero sus menores dispuestos en la esquina izquierda inferior (derecha superior)?

5. Supongamos que para los elementos a_{ij} de la matriz A se verifican las correlaciones

$$a_{ij} = 0, \qquad k < j - i, \quad j - i < l, \tag{93.11}$$

para ciertos números $l \leqslant k$. Tal matriz se llama *matriz de cinta*. Demuéstrese que si para una matriz de cinta A tiene lugar la descomposición (93.1), entonces

$$l_{ij} = 0, \qquad j - i < l, \qquad u_{ij} = 0, \qquad j - i > k.$$

6. Una matriz A se llama *tridiagonal*, si satisface las condiciones (93.11) cuando $k = 1$, $l = -1$. ¿Qué forma tienen las fórmulas (93.5), (93.7) para la matriz tridiagonal?

7. Una matriz A se llama *casi triangular derecha* (izquierda), si satisface las condiciones (93.11) para $k = n$, $l = -1$ ($k = 1$, $l = -n$). ¿Qué forma tienen las fórmulas (93.3) para las matrices casi triangulares?

8. ¿Qué cantidad de operaciones aritméticas debe realizarse para diferentes tipos de matrices cuando se obtienen las descomposiciones del tipo (93.1)?

9. ¿Cómo deben aplicarse las descomposiciones (93.1), (93.5), (93.6) para la resolución de los sistemas de ecuaciones algebraicas lineales?

§ 94. Formas bilineales simétricas

Al examinar las formas bilineales y cuadráticas prestábamos más de una vez especial atención tanto a las formas bilineales simétricas como a las bilineales que engendran formas cuadráticas reales. Sólo dos tipos de las formas

bilineales satisfacen simultáneamente ambas condiciones: la forma bilineal simétrica real y la forma bilineal simétrica hermitiana. Las matrices de estas formas son en cualquier base simétrica real o hermitiana, respectivamente. Ambos tipos de las matrices se reducen, por medio de una transformación congruente, a la forma normal real diagonal.

Según hemos visto, una misma matriz puede reducirse a la forma canónica mediante diferentes transformaciones congruentes. Por eso, en el caso general, la forma canónica no es unívocamente definida. Surge entonces una cuestión ¿qué tienen en común las diferentes formas canónicas a las cuales se reduce una misma matriz? Se sabe que el rango de una matriz no depende de la transformación, razón por la cual, cualquiera que sea el procedimiento de reducción a la forma canónica, el número de las últimas filas nulas será el mismo. Para las matrices simétrica real y hermitiana podemos decir mucho más. La forma canónica de estas matrices puede ser descrita por el número de sus términos positivos y negativos. Tiene lugar el siguiente teorema importante

TEOREMA 94.1 (*ley de inercia de las formas cuadráticas*). *El número de los términos positivos y el de los términos negativos en la forma canónica de una matriz simétrica real, en la transformación congruente ordinaria, y de una matriz hermitiana, en la transformación congruente hermitiana, no dependen de cómo se realiza la reducción.*

DEMOSTRACIÓN. Supongamos que una matriz A satisface las condiciones del teorema. Consideremos la forma cuadrática F con la matriz A de rango r de las variables $x_1, x_2, \ldots, x_n$ y supongamos que dicha matriz se ha reducido a la forma normal mediante dos procedimientos

$$F = y_1^2 + y_2^2 + \ldots + y_k^2 - y_{k+1}^2 - y_{k+2}^2 - \ldots - y_r^2 = \\ = z_1^2 + z_2^2 + \ldots + z_l^2 - z_{l+1}^2 - z_{l+2}^2 - \ldots - z_r^2. \qquad (94.1)$$

Puesto que el paso de las variables $x_1, x_2, \ldots, x_n$ a las variables $y_1, y_2, \ldots, y_n$ se ha realizado mediante una transformación lineal regular, las segundas variables se expresarán linealmente en términos de las primeras, con la particularidad de que el determinante de la matriz de la transformación inversa será diferente de cero. Así pues,

$$y_i = \sum_{s=1}^{n} b_{is} x_s, \qquad \det\begin{pmatrix} b_{11} & \ldots & b_{1n} \\ \cdot & \cdot & \cdot \\ b_{n1} & \ldots & b_{nn} \end{pmatrix} \neq 0. \qquad (94.2)$$

Análogamente

$$z_j = \sum_{t=1}^{n} a_{jt} x_t, \qquad \det\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \cdot & \cdot & \cdot \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \neq 0. \qquad (94.3)$$

Supongamos $k < l$ y escribamos un sistema de ecuaciones

$$y_1 = y_2 = \ldots = y_k = z_{l+1} = \ldots = z_n = 0. \qquad (94.4)$$

Si los primeros miembros de estas igualdades se sustituyen por sus expresiones de (94.2), (94.3), se obtiene un sistema de $n - l + k$ ecuaciones lineales homogéneas con n incógnitas $x_1, x_2, \ldots, x_n$. El número de ecuaciones en este sistema es menor que el número de incógnitas, a consecuencia de lo cual la solución del sistema es real no nula $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Sustituyamos ahora, en la igualdad (94.1), todas las variables por sus expresiones de (94.2), (94.3) y a continuación en lugar de las variables $x_1, x_2, \ldots, x_n$ escribamos los números $\alpha_1, \alpha_2, \ldots, \alpha_n$. Si, realizando tal sustitución, designamos, para brevedad, con $y_i(\alpha)$ y $z_j(\alpha)$ los valores de las variables y_i, z_j, entonces, teniendo en cuenta (94.4), la correlación (94.1) se convierte en la igualdad

$$-y_{k+1}^2(\alpha) - \ldots - y_r^2(\alpha) = z_1^2(\alpha) + \ldots + z_l^2(\alpha).$$

De aquí se desprende

$$z_1(\alpha) = \ldots = z_l(\alpha) = 0. \qquad (94.5)$$

Por otra parte, por la propia elección de los números $\alpha_1, \alpha_2, \ldots, \alpha_n$ tenemos

$$z_{l+1}(\alpha) = \ldots = z_r(\alpha) = \ldots = z_n(\alpha) = 0. \qquad (94.6)$$

De este modo, el sistema de n ecuaciones lineales homogéneas

$$z_i = 0, \qquad i = 1, 2, \ldots, n,$$

con n incógnitas $x_1, x_2, \ldots, x_n$ tiene, en virtud de (94.5), (94.6), una solución no nula $\alpha_1, \alpha_2, \ldots, \alpha_n$, es decir, el determinante de este sistema debe ser igual a cero. Esto contradice a (94.3). A una contradicción análoga llegaremos al suponer $l < k$. Por consiguiente, $l = k$ y el teorema queda demostrado.

Toda forma cuadrática real ordinaria (hermitiana) en un espacio lineal real (complejo) tiene en cualquier base una única matriz simétrica real (hermitiana compleja). Estas matrices satisfacen las condiciones del teorema 94.1. Cualquiera que sea la base el número de términos positivos y negativos de la forma canónica de una matriz es un invariante para la forma cuadrática y se denominan, respectivamente, su *índice de inercia* positivo y negativo. La diferencia entre el índice positivo y el negativo se denomina *signatura* de la forma cuadrática. Ahora se pueden enunciar algunos corolarios útiles del teorema 94.1.

COROLARIO. *Una forma cuadrática es definida positiva (negativa) cuando, y sólo cuando, el índice positivo (negativo) de inercia es igual a n.*

COROLARIO. *Una forma cuadrática es de signo constante cuando, y sólo cuando, uno de los índices de inercia es igual a cero.*

La ley de inercia permite dar cierta clasificación de las formas cuadráticas reales. Llamamos dos formas cuadráticas *equivalentes afines*, si para cada una de ellas se puede elegir tal base que las matrices de dichas formas cuadráticas se hagan iguales. En este caso diremos también que con ayuda de una transformación regular una forma cuadrática se convierte en la otra. Es fácil comprobar que la equivalencia afín de las formas cuadráticas es una relación de equivalencia y dos formas cuadráticas son equivalentes cuando, y sólo cuando, sus matrices en una misma base son congruentes. Por eso, de la ley de inercia proviene que todas las formas cuadráticas reales en el espacio lineal K_n pueden dividirse en clases disjuntas en cada una de las cuales entran formas cuadráticas equivalentes afines y sólo ellas. La clase se caracteriza por el rango y la signatura. La citada división en clases se denomina *clasificación afín* de las formas cuadráticas reales.

Para cualquier rango prefijado r de las formas cuadráticas en una clasificación dada siempre existen dos clases «extremas», las de signaturas $+r$ y $-r$. La primera clase consta de todas las formas cuadráticas no negativas de rango r. A la segunda clase pertenecen todas las formas cuadráticas no positivas de rango r. Ambas clases juntas contienen todas las formas cuadráticas de rango r de signo constante y sólo éstas.

La capacidad de una forma cuadrática de conservar su signo constante se establece, generalmente, por su reducción a la forma canónica, sirviéndose de uno de los métodos descritos. Sin embargo, en algunos casos son de gran interés los criterios inmediatos de la capacidad de conservar los signos constantes. Al tomar en consideración la gran importancia que tienen precisamente tales formas cuadráticas, realizaremos para éstas unas investigaciones adicionales, limitándonos, principalmente, a la consideración de las formas cuadráticas en un espacio real. Esto vez también consideraremos que la matriz de una forma cuadrática es simétrica real. Para el caso de un espacio complejo los resultados serán los mismos, mientras que las demostraciones se diferenciarán en pequeños detalles.

TEOREMA 144 (*Criterio de Sylvester*). *Para que una forma cuadrática sea definida positiva, es necesario y suficiente que todos los menores principales de la matriz de esta forma sean positivos.*

DEMOSTRACIÓN. NECESIDAD. Supongamos que una forma cuadrática con la matriz A es definida positiva. En este caso existe una transformación regular con la matriz P que reduce la forma a una suma de cuadrados. Conforme a (91.9), esto significa que $E = P'AP$

o bien $A = (P^{-1})' P^{-1}$. Haciendo uso de la fórmula Binet—Cauchy obtenemos

$$A\begin{pmatrix}1 & 2 \dots s\\ 1 & 2\dots s\end{pmatrix} = \sum_{1\le k_1<k_2<\dots<k_s\le n} (P^{-1})' \times$$

$$\times \begin{pmatrix}1 & 2 \dots s\\ k_1 & k_2 \dots k_s\end{pmatrix} P^{-1}\begin{pmatrix}k_1 & k_2 \dots k_s\\ 1 & 2 \dots s\end{pmatrix} =$$

$$= \sum_{1\le k_1<k_2<\dots<k_s\le n} \left(P^{-1}\begin{pmatrix}k_1 & k_2 \dots k_s\\ 1 & 2 \dots s\end{pmatrix}\right)^2.$$

Como la matriz P es regular, en las primeras s columnas hay al menos un vector distinto de cero. Por consiguiente, para todo s el segundo miembro de la igualdad obtenida es positivo.

SUFICIENCIA. Supongamos ahora que todos los menores principales de la matriz A de cierta forma cuadrática son positivos. Con ayuda de una transformación, determinada por las fórmulas (93.7), reduzcamos esta forma al tipo canónico. De acuerdo con la hipótesis del teorema y las fórmulas (93.4), todos los coeficientes del tipo canónico serán positivos, es decir, la forma cuadrática es definida positiva.

COROLARIO. *Para que una forma cuadrática sea definida negativa, es necesario y suficiente que todos los menores principales de orden impar sean negativos y todos los menores de orden par, positivos.*

La demostración se deduce del criterio de Sylvester y del hecho de que si A es una matriz de la forma cuadrática definida negativa, entonces $-A$ será una matriz de la forma cuadrática definida positiva.

TEOREMA 94.2 (*criterio de Jacobi*). *Para que una forma cuadrática sea definida positiva, es necesario y suficiente que todos los coeficientes del polinomio característico de la matriz de la forma sean distintos de cero y tengan signos alternados.*

DEMOSTRACIÓN. NECESIDAD. Según se ha observado, por medio de una transformación de las variables con una matriz ortogonal, una forma cuadrática dada puede ser reducida al tipo canónico cuyos coeficientes serán los valores propios $\lambda_1, \lambda_2, \dots, \lambda_n$ de la matriz de la forma. De acuerdo con las condiciones del teorema, los valores propios deben ser positivos. El polinomio característico $f(\lambda)$ es igual a

$$f(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\dots(\lambda - \lambda_n) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots$$

$$\dots + a_1\lambda + a_0$$

siendo todos los coeficientes suyos no nulos con signos alternados, lo que se desprende directamente de las fórmulas de Viète para los coeficientes a_i.

SUFICIENCIA. Supongamos que los coeficientes del polinomio característico son distintos de cero y tienen signos alternados. Las raíces de este polinomio, como valores propios de la matriz simétrica, serán reales y nos resta por señalar que son positivas. Supondremos que esta afirmación está demostrada para todos los polinomios de grado $n - 1$. Ya que todos los coeficientes $f'(\lambda)$ son distintos de cero y tienen signos alternados, entonces, por hipótesis, $f'(\lambda)$ tiene $n - 1$ raíces positivas. Del curso del análisis matemático se conoce que si un polinomio sólo tiene raíces reales, éstas se dividen por las raíces de la derivada. Por eso $f(\lambda)$ tiene por lo menos $n - 1$ raíces positivas. La última raíz será también positiva debido a que es positivo el producto de todas las raíces.

Los criterios para las formas cuadráticas no negativas y no positivas son mucho más complejos y esto se debe, en lo principal, a que en estos casos las matrices son degeneradas. Uno de los métodos que se usa para investigar el carácter constante de los signos de una forma cuadrática está vinculado con la reducción de su matriz a la forma simétrica (93.8), (93.9) como también con el estudio de esta última forma. En virtud de que las matrices de signo constante están estrechamente enlazadas entre sí, nos limitaremos a la consideración de las matrices no negativas.

Llamemos a la matriz H *matriz de conmutaciones*, si cada una de sus filas y cada columna contienen sólo un elemento no nulo y todos los elementos no nulos son iguales a la unidad. Evidentemente, al multiplicar una matriz arbitraria A a la derecha por la matriz de conmutaciones H, en la matriz A se cambian entre sí sus columnas y al multiplicar a la izquierda, se cambian entre sí las filas.

LEMA 96.1. *Para una matriz regular arbitraria A existe tal matriz de conmutaciones H que en la matriz AH todos los menores principales son distintos de cero.*

DEMOSTRACIÓN. A es una matriz regular. Por consiguiente, su primera fila contiene por lo menos un elemento no nulo. Al colocar la columna correspondiente en lugar de la primera, haremos no nulo el menor principal de primer orden. Supongamos que mediante la permutación de las columnas hemos logrado que todos los menores principales de orden hasta k no son iguales a cero. Ahora, si al permutar las últimas $n - k$ columnas no se puede conseguir que el menor principal de orden $(k + 1)$ sea no nulo, esto significa que en las primeras $k + 1$ filas de la matriz A no existe un menor no nulo de orden $k + 1$, es decir, la matriz A debe ser degenerada. La contradicción con la hipótesis del lema significa que el último está demostrado.

TEOREMA 96.6. *Para que una forma cuadrática de rango r con la matriz A sea no negativa, es necesario y suficiente que exista una matriz de conmutaciones H tal, para la cual en la matriz $H'AH$ los primeros r menores principales sean positivos.*

DEMOSTRACIÓN. NECESIDAD. Supongamos que la forma cuadrática de rango r con la matriz A es no negativa. En este caso existe una matriz regular P tal que $A = (P^{-1})'E_rP^{-1}$, donde E_r es una matriz diagonal cuyos r primeros elementos son iguales a uno y los demás, a cero. De acuerdo con el lema 94.1, existe una matriz de conmutaciones H, para la cual todos los menores principales de la matriz $P^{-1}H$ son distintos de cero.

Haciendo uso de la fórmula de Binet—Cauchy, encontramos para $1 \leqslant s < r$

$$H'AH\begin{pmatrix}1 & 2 & \dots & s\\ 1 & 2 & \dots & s\end{pmatrix} =$$

$$= \sum_{1\leqslant k_1<k_2<\dots<k_s\leqslant n} (P^{-1}H)'\begin{pmatrix}1 & 2 & \dots & s\\ k_1 & k_2 & \dots & k_s\end{pmatrix} \times$$

$$\times (E_rP^{-1}H)\begin{pmatrix}k_1 & k_2 & \dots & k_s\\ 1 & 2 & \dots & s\end{pmatrix} =$$

$$= \sum_{1\leqslant k_1<k_2<\dots<k_s\leqslant r} \left\{(P^{-1}H)\begin{pmatrix}k_1 & k_2 & \dots & k_s\\ 1 & 2 & \dots & s\end{pmatrix}\right\}^2 \geqslant$$

$$\geqslant \left\{(P^{-1}H)\begin{pmatrix}1 & 2 & \dots & s\\ 1 & 2 & \dots & s\end{pmatrix}\right\}^2 > 0.$$

SUFICIENCIA. Supongamos que para la forma cuadrática de rango r con la matriz A existe una matriz de conmutaciones H tal que en la matriz $H'AH$ los primeros r menores principales son positivos. La matriz $H'AH$ puede reducirse, de acuerdo con el teorema 93.1, al tipo canónico mediante la transformación con una matriz triangular. De conformidad con (93.4), los coeficientes no nulos del tipo canónico de la matriz $H'AH$ y, por tanto, de la matriz A serán positivos, es decir, la forma cuadrática es no negativa.

En cuanto a las formas cuadráticas de signo no constante, conviene decir que para ellas no existen análogos completos de los teoremas 94.1, 94.4. Sólo tiene lugar el

TEOREMA 11.5. *Si una forma cuadrática tiene una matriz simétrica A del tipo* (93.8), (93.9), *sus índices de inercia coinciden con los de inercia para la forma cuadrática «truncada» que se determina mediante la matriz B de* (93.8).

DEMOSTRACIÓN. De acuerdo con el teorema 93.1, la matriz A puede ser reducida a la forma canónica mediante la transformación con una matriz triangular derecha, con la particularidad de que para los coeficientes no nulos del tipo canónico se verifican las correlaciones (93.4). Para la matriz B de la forma cuadrática «truncada» se satisface también las condiciones del teorema 93.1 y para los coefi-

cientes de su tipo canónico se verifican asimismo las correlaciones (93.4). Por eso, los índices de inercia de las formas cuadráticas, determinados por las matrices A y B, coinciden.

El singular interés respecto a las formas cuadráticas de signo constante se debe al gran campo de sus aplicaciones. Una de las aplicaciones más importantes es la introducción de una métrica en un espacio lineal. Toda forma bilineal, polar respecto a cierta forma cuadrática definida positiva, puede considerarse como producto escalar y, por lo tanto, con su ayuda podemos convertir un espacio lineal en un espacio euclídeo o unitario. El cumplimiento de los axiomas para estos espacios es obvio. Una significación no menos importante para introducir una métrica, especialmente en los subespacios, tienen también las formas no negativas. Como ejemplo de la utilización del carácter constante de los signos demostremos que es válido el

TEOREMA 94.1. *Para que una forma bilineal hermitiana regular sea reducible a la forma diagonal, es suficiente que su parte simétrica (o antisimétrica) sea estrictamente de signo constante.*

DEMOSTRACIÓN. Consideraremos el caso en que la parte simétrica es definida positiva. Sea A una matriz de la forma bilineal, entonces $\frac{1}{2}(A+A^*)$ será la matriz de la parte simétrica. Puesto que la parte simétrica es definida positiva, la matriz $\frac{1}{2}(A+A^*)$ es congruente de la matriz unidad según Hermite. Por consiguiente, existe una matriz regular S tal que

$$\frac{1}{2}S'(A+A')\bar{S}=E. \qquad (94.7)$$

Mostramos que la matriz $S'A\bar{S}$ es normal. De (94.7) tenemos

$$\bar{S}S'=2(A+A^*)^{-1}.$$

Por ello

$$\begin{aligned}(S'A\bar{S})(S'A\bar{S})^*-(S'A\bar{S})^*(S'A\bar{S})&=S'(A\bar{S}S'A^*-A^*\bar{S}S'A)=\\ &=2S'(A(A+A^*)^{-1}A^*-A^*(A+A^*)^{-1}A)\bar{S}=\\ &=2S'((A^{*-1}(A+A^*)A^{-1})^{-1}-(A^{-1}(A+A^*)A^{*-1})^{-1})\bar{S}=\\ &=2S'((A^{*-1}+A^{-1})^{-1}-(A^{*-1}+A^{-1})^{-1})\bar{S}=0.\end{aligned}$$

Siendo normal, la matriz $S'AS$ se reduce a la forma diagonal mediante la transformación congruente según Hermite con una matriz unitaria.

De esta modo, la matriz A de la forma bilineal hermitiana es congruente según Hermite de la matriz diagonal, lo que se trataba de demostrar. Todos los casos restantes se examinan de modo análogo.

Ejercicios.

1. Demuéstrese que si todos los menores principales de una matriz simétrica real o hermitiana compleja son distintos de cero, entonces el número de sus valores propios positivos y negativos coincide respectivamente con el número de los términos positivos y negativos de la sucesión (93.4).

2. Demuéstrese que si una matriz es definida positiva, todo menor diagonal es positivo.

3. Demuéstrese que una matriz simétrica de rango r siempre tiene un menor diagonal, aunque sea único, de rango r, distinto de cero.

4. Demuéstrese que el elemento máximo de una matriz definida positiva se encuentra en la diagonal principal.

5. Demuéstrese que la matriz A es definida positiva, si para todo i

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|.$$

6. Demuéstrese que para toda matriz simétrica A de rango r existe una matriz de conmutaciones H tal que entre los primeros r menores principales de la matriz $H'AH$ no hay dos adyacentes que sean nulos, con la particularidad de que el menor de r-ésimo orden es distinto de cero.

7. Demuéstrese que la matriz $H'AH$ del ejercicio 6 puede ser representada en la forma $H'AH = S'DS$, donde S es una matriz triangular derecha con elementos diagonales iguales a uno, D es una matriz diagonal celular cuyas células son de primero y segundo órdenes.

8. Demuéstrese que toda matriz no negativa de rango r puede ser representada como la suma de r matrices no negativas de rango 1.

9. Sean A, B las matrices definidas positivas con los elementos a_{ij}, b_{ij}. Demuéstrese que la matriz C con los elementos $c_{ij} = a_{ij}b_{ij}$ es también definida positiva.

§ 95. Hipersuperficies de segundo grado

Con el estudio de las formas cuadráticas reales está estrechamente relacionada la investigación de otros objetos, a saber, hipersuperficies de segundo grado. Deseando subrayar el carácter geométrico de muchas propiedades de las hipersuperficies, en adelante llamaremos a los vectores casi siempre puntos del espacio $\mathbf{R}_n$.

Se denomina *hipersuperficie* ***f*** *de segundo grado* en el espacio $\mathbf{R}_n$ un conjunto de puntos cuyas coordenadas $x_1, x_2, \ldots, x_n$ satisfacen la ecuación

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji}x_i x_j - 2\sum_{k=1}^{n} b_k x_k + c = 0, \tag{95.1}$$

donde a_{ji}, b_i, c son unos números reales.

Simplifiquemos la anotación. Al igual que en el caso de las formas cuadráticas, supondremos que la matriz A con los coeficientes a_{ji} es simétrica. Designemos con b un vector con las coordenadas $b_1, b_2, \ldots, b_n$. Introduzcamos en el espacio $\mathbf{R}_n$ un producto escalar como suma de productos de las coordenadas, tomados dos a dos.

Ahora, la hipersuperficie f de segundo grado en el espacio R_n puede considerarse como un conjunto de los puntos x del espacio euclídeo R_n que satisfacen la ecuación

$$(Ax, x) - 2(b, x) + c = 0 \tag{95.2}$$

o bien, por ser la matriz A simétrica, la ecuación

$$(x, Ax) - 2(b, x) + c = 0.$$

La investigación de las hipersuperficies de segundo grado la empezaremos con el estudio de la disposición conjunta de estas superficies y de líneas rectas. Tomamos una recta arbitraria en el espacio R_n. Supongamos que esta recta pasa por el punto x_0 y tiene un vector director l. Los puntos x de dicha recta se definen mediante la igualdad

$$x = x_0 + lt \tag{95.3}$$

para cualesquiera números reales t. Al sustituir la expresión dada para x en (95.2), obtendremos

$$t^2(Al, l) - 2t\{(b, l) - (Al, x_0)\} + \\ + (Ax_0, x_0) - 2(b, x_0) + c = 0. \tag{95.4}$$

De este modo, los puntos de intersección de la recta (95.3) con la hipersuperficie (95.2) se determinan por las raíces de la ecuación cuadrática (95.4).

Diremos que la recta (95.3) con el vector director l es de dirección *no asintótica* (*asintótica*) respecto de la hipersuperficie (95.2), si $(Al, l) \neq 0$ ($(Al, l) = 0$).

Consideraremos una recta cualquiera que tiene la dirección no asintótica l y atraviesa la hipersuperficie. Los puntos de intersección determinan en cada una de estas rectas un segmento el cual llamaremos, por analogía con la geometría elemental, *cuerda*. Designemos con L el conjunto de puntos medios de todas las cuerdas. Si los extremos de una cuerda son contraídos a un punto, éste se considerará también como punto medio de la cuerda. Probemos que L pertenece a cierta hipersuperficie.

Los extremos de cualquier cuerda se determinan por los valores del parámetro t coincidentes con las raíces de la ecuación (95.4). Por ello el punto medio de la cuerda se determina por el valor de t igual a la semisuma de las raíces. De conformidad con las fórmulas de Viète esto nos da

$$t = \frac{(b, l) - (Al, x_0)}{(Al, l)}. \tag{95.5}$$

Si z_0 es el punto medio de la cuerda, entonces

$$z_0 = x_0 + \frac{(b, l) - (Al, x_0)}{(Al, l)}\, l.$$

Ahora tenemos

$$(Al,\ z_0) = \left(Al,\ x_0 + \frac{(b,\ l)-(Al,\ x_0)}{(Al,\ l)}\, l\right) =$$

$$= (Al,\ x_0) + \frac{(b,\ l)-(Al,\ x_0)}{(Al,\ l)}\,(Al,\ l) = (b,\ l).$$

Así pues, los puntos medios de todas las cuerdas satisfacen la ecuación

$$(Al,\ x) = (b,\ l). \tag{95.6}$$

Como el segundo miembro de la ecuación no depende de x_0, entonces, de acuerdo con la fórmula (46.8), esta ecuación determina un hiperplano cuyo vector normal es igual a Al.

El hiperplano (95.6) se llama *hiperplano diametral conjugado* de la dirección l respecto a la hipersuperficie (95.2).

La forma explícita de la ecuación del hiperplano diametral permite establecer toda una serie de propiedades importantes que poseen las hipersuperficies de segundo grado. Sea A una matriz regular. Entonces, para cualesquiera vectores linealmente independientes $l_1, l_2, \ldots, l_n$ serán también linealmente independientes los vectores $Al_1, Al_2, \ldots, Al_n$. Supondremos luego que todas las direcciones $l_1, l_2, \ldots, l_n$ son no asintóticas. Esto tendrá lugar a ciencia cierta en el caso, por ejemplo, cuando la forma cuadrática $(Ax,\ x)$ es definida positiva. Por consiguiente, se puede construir un sistema de n hiperplanos diametrales conjugados de las direcciones $l_1, l_2, \ldots, l_n$. Los hiperplanos tendrán un único punto común x^*. Ahora, de la fórmula (95.6) se desprenden las igualdades

$$(Ax^* - b,\ l_i) = 0$$

para $i = 1, 2, \ldots, n$. En virtud de la independencia lineal de los vectores l_i, esto significa que $Ax^* - b = 0$, es decir, el punto x^* no es otra cosa que la solución del sistema de ecuaciones algebraicas lineales

$$Ax = b. \tag{95.7}$$

La solución del sistema con una matriz regular es única, razón por la cual el punto construido x^* no depende, en realidad, de cómo se escogen los vectores $l_1, l_2, \ldots, l_n$.

Unos cálculos simples muestran que para todo punto x^* es válida la correlación

$$(Ax,\ x) - 2(b,\ x) + c = (A(x - x^*),\ x - x^*) + $$
$$+ 2(Ax^* - b,\ x - x^*) + (Ax^*,\ x^*) - 2(b,\ x^*) + c. \tag{95.8}$$

Si, en cambio, x^* es la solución del sistema (95.7), entonces respecto de tal punto la hipersuperficie (95.2) posee la propiedad importante

de simetría. A saber, cualquiera que sea x, el primer miembro de (95.2) toma valores iguales en los puntos

$$x = x^* + (x - x^*), \qquad x' = x^* - (x - x^*). \qquad (95.9)$$

De aquí se deduce, en particular, que ambos puntos, x, x', se ubican o no se ubican en la hipersuperficie (95.2) simultáneamente. La igualdad

$$x^* = \frac{1}{2}(x + x')$$

permite llamar el punto x^* *centro de simetría* de la hipersuperficie. Si en la hipersuperficie (94.2) se dispone aunque sea un solo punto de R_n, el centro de simetría se denomina *real*. En el caso contrario se llama *imaginario*.

Sea, ahora, x^* un centro de simetría, es decir, para todo x el primer miembro de (95.2) toma valores iguales en los puntos x, x'. Por consiguiente,

$$(Ax, x) - 2(b, x) + c = (Ax', x') - 2(b, x') + c.$$

De acuerdo con (95.8), (95.9), esto es posible sólo en el caso de que para cualquier x

$$(Ax^* - b, x - x^*) = 0.$$

Mas, la última identidad es válida cuando, y sólo cuando, $Ax^* - b = 0$, es decir, cuando el punto x^* sea la solución del sistema (95.7). Cabe señalar que aquí nunca suponíamos la regularidad de la matriz A, ni tampoco le presumíamos otras singularidades, salvo la simetría. Por esta razón:

Para que el sistema $Ax = b$ tenga solución, es necesario y suficiente que la hipersuperficie (95.2) cuente con un centro de simetría. El conjunto de todas las soluciones coincide con el conjunto de todos los centros de simetría.

De este modo, se pone de manifiesto una relación muy profunda entre los sistemas de ecuaciones algebraicas lineales y las hipersuperficies de segundo grado. Esta relación se utiliza ampliamente al construir los más diversos algoritmos de cálculo. En particular, en la construcción del sistema de hipersuperficies diametrales se basa un gran grupo de métodos que forma parte del grupo de los llamados métodos de direcciones conjugadas. Estos métodos se tratarán en el último capítulo de la obra.

En el caso general la investigación de las hipersuperficies de segundo grado puede fundarse en la reducción de ellas a la forma canónica, casi por analogía completa con las formas cuadráticas. Pero en este caso, además de las transformaciones regulares lineales de las variables, se exigirán las operaciones de desplazamiento.

Consideremos una transformación cualquiera de las variables $x = Py$ que reduce la forma cuadrática (Ax, x) a la forma normal. En términos de las variables $y_1, y_2, \ldots, y_n$ la ecuación de la hipersuperficie tendrá por expresión

$$y_1^2 + \ldots + y_k^2 - y_{k+1}^2 - \ldots - y_r^2 - \\ - 2d_1y_1 - \ldots - 2d_ry_r - 2d_{r+1}y_{r+1} - \ldots - 2d_ny_n + c = 0.$$

Realicemos ahora el desplazamiento de las variables de acuerdo con las fórmulas

$$z_i = \begin{cases} y_i - d_i, & 1 \leqslant i \leqslant k, \\ y_i + d_i, & k+1 \leqslant i \leqslant r, \\ y_i & r+1 \leqslant i \leqslant n. \end{cases}$$

En términos de estas variables la ecuación toma la forma

$$z_1^2 + \ldots + z_k^2 - z_{k+1}^2 - \ldots - z_r^2 - 2d_{r+1}z_{r+1} - \ldots - 2d_nz_n + p = 0.$$

Supongamos que uno de los números $d_{r+1}, \ldots, d_n$, por ejemplo d_n, es distinto de cero. Hagamos

$$v_i = \begin{cases} z_i, & i < n, \\ d_{r+1}z_{r+1} + \ldots + d_nz_n, & i = n, \end{cases}$$

y a continuación realicemos nuevamente un desplazamiento

$$u_i = \begin{cases} v_i, & i < n \\ v_i - \frac{p}{2} & i = n. \end{cases}$$

Ahora, la ecuación de la hipersuperficie adquiere la forma:

$$\pm u_1^2 \pm u_2^2 \pm \ldots \pm u_r^2 - 2u_n = 0, \quad 1 \leqslant r \leqslant n - 1. \quad (95.10)$$

Si entre los números $d_{r+1}, \ldots, d_n, p$ no hay ninguno que sea igual a cero, la ecuación de la hipersuperficie toma la forma siguiente:

$$\pm u_1^2 \pm u_2^2 \pm \ldots \pm u_r^2 = 0, \quad 1 \leqslant r \leqslant n. \quad (95.11)$$

Y, por fin, si los números $d_{r+1}, \ldots, d_n$ son nulos, mientras que $p \neq 0$, entonces, al poner $u_i = z_i/|p|^{1/2}$ para todo i, obtendremos una forma más para la ecuación de la hipersuperficie. A saber,

$$\pm u_1^2 \pm u_2^2 \pm \ldots \pm u_r^2 \pm 1 = 0, \quad 1 \leqslant r \leqslant n. \quad (95.12)$$

En virtud de la ley de inercia de las formas cuadráticas, las superficies definidas por diversas ecuaciones del tipo (95.10)—(95.12) no pueden convertirse la una en la otra con ayuda de la transformación lineal de las variables y un desplazamiento. En este caso

deben considerarse diferentes las ecuaciones que no pueden ser transformadas la una en la otra multiplicándolas por (-1) y cambiando la numeración de las coordenadas. Igual que en el caso de las formas cuadráticas, esta vez obtuvimos también una partición de todas las hipersuperficies de segundo grado en clases disjuntas.

Al reducir las hipersuperficies de segundo grado a la forma canónica se utilizan con frecuencia sólo las operaciones de traslado y transformaciones lineales de las variables con matrices ortogonales. Esto se debe, principalmente, a que ambos tipos de las transformaciones indicadas no cambian las distancias entre los puntos. En este caso las formas canónicas serán un tanto diferentes, aunque en total se obtienen del mismo modo que las anteriores. Por ejemplo, en el caso del espacio R_2, la hipersuperficie de segundo grado puede ser reducida sólo a uno de los siguientes tipos:

$$
\begin{aligned}
&\text{I.}\ \ \lambda_1 x^2 + \lambda_2 y^2 + a_3 = 0,\\
&\text{II.}\ \ \lambda_2 y^2 + b_0 x = 0, \qquad\qquad (95.13)\\
&\text{III.}\ \ \lambda_1 x^2 + a_3 = 0,
\end{aligned}
$$

y en el caso del espacio R_3, a uno de los siguientes tipos:

$$
\begin{aligned}
&\text{I.}\ \ \lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + a_4 = 0,\\
&\text{II.}\ \ \lambda_1 x^2 + \lambda_2 y^2 + b_0 z = 0,\\
&\text{III.}\ \ \lambda_1 x^2 + \lambda_2 y^2 + a_4 = 0, \qquad\qquad (95.14)\\
&\text{IV.}\ \ \lambda_1 y^2 + b_0 x = 0,\\
&\text{V.}\ \ \lambda_1 x^2 + a_4 = 0.
\end{aligned}
$$

En todas las ecuaciones (95.13), (95.14) los coeficientes de las variables escritas son distintos de cero. El término independiente puede ser igual a cero. De acuerdo con la terminología aceptada, las hipersuperficies en el espacio R_2 se llamarán *líneas de segundo orden* y en el espacio R_3, *superficies de segundo grado*. Tomando en consideración los intereses de los diferentes apartados de las matemáticas, estudiaremos más detalladamente las líneas y las superficies de segundo grado según sus formas canónicas (95.13), (95.14).

Ejercicios.

1. Sea A una matriz definida positiva. Demuéstrese que en la solución del sistema $Ax = b$ la expresión $(Ax, x) - 2(b, x)$ alcanza su valor mínimo.

2. Sea A una matriz definida positiva. Demuéstrese que en la recta (95.3) la expresión $(Ax, x) - 2(b, x)$ alcanza su valor mínimo para el valor de t tomado de (95.5).

3. Demuéstrese que para que una dirección cualquiera sea no asintótica para la hipersuperficie (95.2), es necesario y suficiente que la forma cuadrática (Ax, x) sea definida positiva o negativa.

4. ¿Qué propiedad de simetría posee un hiperplano diametral conjugado de la dirección l, si l es un vector propio de la matriz A, correspondiente al valor propio no nulo?

5. Demuéstrese que el sistema $Ax = b$ no tiene solución cuando, y sólo cuando, el hiperplano (95.2) se reduce a la forma canónica (95.10).

§ 96. Líneas de segundo orden

Estudiaremos las líneas de segundo orden mediante las ecuaciones (95.13). Supongamos que la ecuación de una línea tiene por expresión

$$\lambda_1 x^2 + \lambda_2 y^2 + a_0 = 0. \qquad (96.1)$$

1.1. *El número a_0 no es nulo; los números λ_1, λ_2, son de signos iguales, contrarios al signo de a_0.* Escribamos (96.1) en la forma siguiente

$$\frac{x^2}{\left(\sqrt{-\frac{a_0}{\lambda_1}}\right)^2} + \frac{y^2}{\left(\sqrt{-\frac{a_0}{\lambda_2}}\right)^2} = 1$$

y designamos

$$a = \sqrt{-\frac{a_0}{\lambda_1}}, \quad b = \sqrt{-\frac{a_0}{\lambda_2}} \qquad (96.2)$$

Según la condición, los números a y b son reales, por lo cual la ecuación (96.1) es equivalente a la ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \qquad (96.3)$$

Una línea, descrita por esta ecuación, se denomina *elipse* (fig. 96.1) y la propia ecuación se llama *ecuación canónica de la elipse*. Demos a conocer algunas propiedades de la elipse. Una elipse es una línea acotada. Como se deduce de la ecuación (96.3), para todos los puntos de la elipse se tiene: $|x| \leqslant a$, $|y| \leqslant b$. La elipse cuenta con dos ejes de simetría: el eje Ox y el eje Oy, como también un centro de simetría que es el origen de coordenadas. Esto se deduce del hecho que a la par con un punto de coordenadas (x, y), a la elipse pertenecen los puntos que tienen las coordenadas $(x, -y)$, $(-x, y)$, $(-x, -y)$. Los ejes de simetría se llaman *ejes principales* de la elipse; el centro de simetría es a la vez *centro* de la elipse. Si $a > b$, entonces Ox lleva el nombre de *eje mayor* de la elipse y Oy, *eje menor* de la elipse. Los puntos de intersección de los ejes principales de la elipse con la propia elipse se llaman *vértices* de

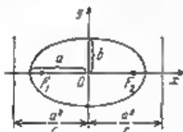


Fig. 96.1.

la elipse. Cuando $a = b$, la elipse se convierte en una circunferencia de radio a y centro en el origen de coordenadas. Supongamos, para concretar, que $a > b$ y designemos

$$c^2 = a^2 - b^2. \tag{96.4}$$

Los puntos F_1, F_2 con las coordenadas $(-c, 0)$, $(+c, 0)$ se denominan *focos* de la elipse.

TEOREMA 96.1. *La suma de las distancias desde cualquier punto de la elipse hasta sus focos es una magnitud constante, igual a 2a.*

DEMOSTRACION. Para todo punto $M(x, y)$ de la elipse tenemos

$$y^2 = b^2 - \frac{b^2x^2}{a^2}.$$

Para este mismo punto calculamos

$$\rho(M, F_2) = ((x-c)^2 + y^2)^{1/2} = \left(x^2 - 2xc + c^2 + b^2 - \frac{b^2x^2}{a^2}\right)^{1/2} =$$
$$= \left(x^2\left(1 - \frac{b^2}{a^2}\right) - 2xc + a^2\right)^{1/2} = \left(\frac{x^2c^2}{a^2} - 2xc + a^2\right)^{1/2} =$$
$$= \left(\left(-\frac{c}{a}x + a\right)^2\right)^{1/2} = -\frac{c}{a}x + a.$$

La última igualdad es válida, puesto que $-\frac{c}{a}x + a > 0$ porque $|x| \leqslant a$ y $c/a < 1$. Luego,

$$\rho(M, F_1) = ((x+c)^2 + y^2)^{1/2} = \left(x^2 + 2xc + c^2 + b^2 - \frac{b^2x^2}{a^2}\right)^{1/2} =$$
$$= \left(x^2\left(1 - \frac{b^2}{a^2}\right) + 2xc + a^2\right)^{1/2} = \left(\frac{x^2c^2}{a^2} + 2xc + a^2\right)^{1/2} =$$
$$= \left(\left(\frac{cx}{a} + a\right)^2\right)^{1/2} = \frac{c}{a}x + a.$$

En definitiva tenemos

$$\rho(M, F_1) + \rho(M, F_2) = -\frac{c}{a}x + a + \frac{c}{a}x + a = 2a.$$

I.2. *El número a_0 no es nulo; los números λ_1, λ_2, a_0 son de signos iguales.* Designemos

$$a = \sqrt{\frac{a_0}{\lambda_1}}, \quad b = \sqrt{\frac{a_0}{\lambda_2}}, \tag{96.5}$$

entonces la ecuación (96.1) es equivalente a la ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1. \tag{96.6}$$

Está claro que no hay ningún punto del plano que satisfaga (96.6). La ecuación (96.6) suele tratarse como la ecuación de una elipse *imaginaria*.

1.3. *El número a_0 es nulo; los números λ_1, λ_2 son de signos iguales.* Designemos

$$a=\sqrt{\frac{1}{|\lambda_1|}},\qquad b=\sqrt{\frac{1}{|\lambda_2|}},$$

en este caso la ecuación (96.1) es equivalente a la siguiente:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0. \tag{96.7}$$

Está claro que solamente el origen de coordenadas satisface la ecuación (96.7). La ecuación (96.7) suele tratarse como la ecuación de una elipse *degenerada*.

1.4. *El número a_0 no es igual a cero; los números λ_1, λ_2 son de signos contrarios.* Por introducción de los nuevos coeficientes análogos a (96.2), (96.5), la ecuación (96.1) se reduce, salvo la redesignación de las variables, a una ecuación equivalente

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1. \tag{96.8}$$

Una línea, descrita por esta ecuación, se llama *hipérbola* (fig. 96.2) y la propia ecuación, ecuación *canónica* de la hipérbola. A diferencia de la elipse, la hipérbola es una línea no acotada. Igual que en el caso de una elipse, los ejes de simetría de la hipérbola son los ejes de coordenadas y el centro de simetría es el origen de coordenadas. Los ejes de simetría se denominan *ejes principales* de la hipérbola; el centro de simetría, *centro* de la hipérbola. Uno de los ejes principales (Ox) se interseca con la hipérbola en dos puntos llamados *vértices* de la hipérbola. Este eje se llama eje *real* de la hipérbola. El otro eje (Oy) no tiene puntos comunes con la hipérbola y se llama, por eso, eje *imaginario* de la hipérbola. Designemos

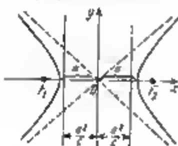


Fig. 96.2.

$$c^2=a^2+b^2.$$

Los puntos F_1, F_2, cuyas coordenadas son $(-c, 0)$, $(+c, 0)$ se denominan *focos* de la hipérbola.

TEOREMA 96.2. *La magnitud absoluta de la diferencia entre las distancias desde cualquier punto de la hipérbola hasta sus focos es constante e igual a $2a$.*

DEMOSTRACIÓN. Para todo punto $M(x, y)$ de la hipérbola se tiene

$$y^2=-b^2+\frac{b^2x^2}{a^2}.$$

Para el mismo punto calculamos

$$\rho(M, F_2) = ((x-c)^2 + y^2)^{1/2} = \left(x^2 - 2xc + c^2 - b^2 + \frac{b^2x^2}{a^2}\right)^{1/2} =$$
$$= \left(x^2\left(1 + \frac{b^2}{a^2}\right) - 2xc + a^2\right)^{1/2} = \left(\frac{x^2c^2}{a^2} - 2xc + a^2\right)^{1/2} =$$
$$= \left(\left(\frac{c}{a}x - a\right)^2\right)^{1/2} = \left|\frac{c}{a}x - a\right|.$$

Luego

$$\rho(M, F_1) = ((x+c)^2 + y^2)^{1/2} =$$
$$= \left(x^2 + 2xc + c^2 + \frac{b^2}{a^2}x^2 - b^2\right)^{1/2} = \left(x^2\left(1 + \frac{b^2}{a^2}\right) + 2xc + a^2\right)^{1/2} =$$
$$= \left(\frac{x^2c^2}{a^2} + 2xc + a^2\right)^{1/2} = \left(\left(\frac{c}{a}x + a\right)^2\right)^{1/2} = \left|\frac{c}{a}x + a\right|.$$

Para todos los puntos de la hipérbola tenemos $|x| \geqslant a$ y $c/a > 1$. Por esto

$$\rho(M, F_2) = \begin{cases} \frac{c}{a}x - a & \text{para toda } x > 0, \\ -\frac{c}{a}x + a & \text{para todo } x < 0, \end{cases}$$

$$\rho(M, F_1) = \begin{cases} \frac{c}{a}x + a & \text{para toda } x > 0, \\ -\frac{c}{a}x - a & \text{para todo } x < 0. \end{cases}$$

En definitiva,

$$|\rho(M, F_1) - \rho(M, F_2)| = 2a.$$

Consideraremos la parte de la hipérbola dispuesta en el primer cuadrante. Para esta parte $x \geqslant a$, $y \geqslant 0$. La ecuación (96.8) en el primer cuadrante es equivalente a la siguiente

$$y = \frac{b}{a}\sqrt{x^2 - a^2},$$

al, naturalmente, se considera que $b > 0$, $a > 0$. Es fácil convencerse de que esta función puede ser representada en la forma siguiente:

$$y = \frac{b}{a}x - \frac{ba}{a + \sqrt{x^2 - a^2}}. \qquad (96.9)$$

A la par con la función (96.9) examinemos la ecuación de una recta

$$y' = \frac{b}{a}x. \qquad (96.10)$$

Designaremos mediante $M(x, y)$ y $M'(x, y')$ los puntos de la hipérbola (96.9) y de la recta (96.10) que tienen una misma abscisa x.

Al crecer ilimitadamente x, la diferencia

$$y'-y=\frac{ba}{x+\sqrt{x^2-a^2}}$$

va decreciendo de manera monótona, quedando positiva, y tiende hacia cero. Por consiguiente, los puntos M y M' van acercándose, pero el punto M de la hipérbola queda siempre por debajo del punto M' en la recta (96.10).

Una propiedad análoga tiene lugar también para las otras partes de la hipérbola. Una de las rectas

$$y=\frac{b}{a}x, \qquad y=-\frac{b}{a}x \qquad (96.11)$$

desempeña el papel de la recta (96.10). Estas rectas se llaman *asíntotas* de la hipérbola.

Observemos que hemos tratado (96.8) como la ecuación de la hipérbola. Sin embargo, del curso escolar se conoce otra ecuación que también se denomina *ecuación de la hipérbola*.

Teniendo presente (96.11), realicemos el siguiente cambio de coordenadas

$$x'=\frac{x}{a}-\frac{y}{b}, \qquad y'=\frac{x}{a}+\frac{y}{b}.$$

De (96.8) tenemos

$$\left(\frac{x}{a}-\frac{y}{b}\right)\left(\frac{x}{a}+\frac{y}{b}\right)=1.$$

Por lo tanto, en el nuevo sistema de coordenadas (no rectangular, en el caso general) la ecuación de la hipérbola tiene la forma

$$x'y'=1 \qquad (96.12)$$

o bien

$$y'=\frac{1}{x'}.$$

Ésta es precisamente la ecuación conocida del curso escolar. La ecuación (96.12) se denomina ecuación de la hipérbola respecto de sus asíntotas.

1.5. *El número a_0 es igual a cero; los números λ_1, λ_2 son de signos contrarios.* Al efectuar el cambio habitual de los coeficientes, obtendremos la ecuación

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0 \qquad (96.13)$$

equivalente a la ecuación (96.1). De esta ecuación obtenemos

$$\left(\frac{x}{a}-\frac{y}{b}\right)\left(\frac{x}{a}+\frac{y}{b}\right)=0$$

o bien

$$y = \frac{b}{a} x, \qquad y = -\frac{b}{a} x. \tag{96.14}$$

De este modo, la ecuación (96.13) es la ecuación de una línea que se descompone en *dos rectas* (96.14) *que se cortan*.

Consideraremos ahora la segunda ecuación de (95.13). Esta tiene la forma

$$\lambda_2 y^2 + b_0 x = 0. \tag{96.15}$$

II.6. *Ambos números* λ_2, b_0 *son distintos de cero.* Designemos

$$2p = -\frac{b_0}{\lambda_2} \neq 0.$$

Ahora, la ecuación (96.15) resulta equivalente a la siguiente:

$$y^2 = 2px. \tag{96.16}$$

La línea descrita por esta ecuación se denomina *parábola* (fig. 96.3) y la propia ecuación lleva el nombre de ecuación *canónica* de la parábola. Sin restringir la generalidad, se puede considerar que $p > 0$, puesto que para $p < 0$ se obtiene una línea simétrica respecto del eje Oy. Análogamente a la hipérbola, la parábola es una línea no acotada. Tiene solamente un eje de simetría, el eje Ox, y no tiene centro de simetría. El punto de intersección del eje de la parábola con la misma parábola se llama *vértice* de la parábola. El punto F cuyas coordenadas son $\left(\frac{p}{2}, 0\right)$ se denomina *foco* de la parábola. La recta L definida por la ecuación

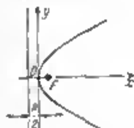


Fig. 96.3.

$$x = -\frac{p}{2} \tag{96.17}$$

se llama directriz de la parábola.

TEOREMA 96.1. *La distancia entre cualquier punto de una parábola y la directriz es igual a la distancia entre dicho punto y el foco de la parábola.*

DEMOSTRACIÓN. Para cualquier punto $M(x, y)$ de la parábola tenemos

$$\rho(L, M) = x + \frac{p}{2},$$

y luego

$$\rho(F, M)=\left(\left(x-\frac{p}{2}\right)^2+y^2\right)^{1/2}=\left(\left(x-\frac{p}{2}\right)^2+2px\right)^{1/2}=$$
$$=\left(x^2-px+\frac{p^2}{4}+2px\right)^{1/2}=\left(x^2+px+\frac{p^2}{4}\right)^{1/2}=$$
$$=\left(\left(x+\frac{p}{2}\right)^2\right)^{1/2}=x+\frac{p}{2}$$

puesto que $x \geqslant 0$ y $p > 0$.

Consideraremos, por fin, la tercera ecuación de (95.13). Es de la forma más sencilla:

$$\lambda_1 x^2 + a_0 = 0. \tag{96.18}$$

III.7. *El número a_0 no es igual a cero; el signo del número λ_1 es contrario al de a_0.* Designemos

$$a^2 = -\frac{a_0}{\lambda_1},$$

entonces la ecuación de la línea (96.18) será equivalente a la ecuación

$$x^2 - a^2 = 0 \tag{96.19}$$

o bien

$$x = a, \qquad x = -a. \tag{96.20}$$

Por consiguiente, la ecuación de la línea (96.19) es la ecuación de una línea que se descompone en dos rectas *paralelas* (96.20).

III.8. *El número a_0 no es igual a cero; el signo del número λ_1 coincide con el de a_0.* Designemos

$$a^2 = \frac{a_0}{\lambda_1},$$

entonces la ecuación (96.18) será equivalente a la ecuación

$$x^2 + a^2 = 0. \tag{96.21}$$

Está claro que no existe ningún punto del plano cuyas coordenadas satisfagan esta ecuación. De (96.21) suele decirse como de una ecuación que define *dos rectas imaginarias*.

III.9. *El número a_0 es igual a cero.* En este caso (96.18) es equivalente a la ecuación

$$x^2 = 0. \tag{96.22}$$

Por analogía con la ecuación (96.19), suele decirse que la ecuación (96.22) define *dos rectas coincidentes*, cada una de las cuales se determina mediante la ecuación

$$x = 0.$$

Ha de señalarse que para todos los puntos de una elipse o una hipérbola tienen lugar las siguientes igualdades:

$$\rho(M, F_2)=\frac{c}{a}\left|x-\frac{a^2}{c}\right|, \qquad (96.23)$$

$$\rho(M, F_1)=\frac{c}{a}\left|x+\frac{a^2}{c}\right|.$$

Las rectas α_i ($i = 1, 2$), definidas por las ecuaciones

$$x-\frac{a^2}{c}=0, \qquad x+\frac{a^2}{c}=0, \qquad (96.24)$$

se llaman *directrices* de la elipse y de la hipérbola. Atribuiremos a la directriz y el foco números idénticos, si se disponen en un mismo semiplano definido por el eje Oy. Ahora podemos mostrar que:

La razón entre las distancias $\rho(M, F_i)$ y $\rho(M, \alpha_i)$ es una magnitud constante para todos los puntos M de la elipse, hipérbola y parábola.

Para la parábola esta afirmación se deduce del teorema 96.3. Para la elipse y la hipérbola, de las ecuaciones (96.23), (96.24). La razón

$$e=\frac{\rho(M, F_i)}{\rho(M, \alpha_i)}$$

se denomina *excentricidad*. Se tiene

para la elipse:

$$e=\frac{c}{a}=\left(1-\frac{b^2}{a^2}\right)^{1/2}<1,$$

para la hipérbola:

$$e=\frac{c}{a}=\left(1+\frac{b^2}{a^2}\right)^{1/2}>1,$$

para la parábola:

$$e=1.$$

Ejercicios.

1. ¿Qué representa en el en hiperplano diametral conjugado de una dirección dada para las líneas de segundo orden?
2. Escríbanse las ecuaciones de una línea tangente para una elipse, una hipérbola y una parábola.
3. Demuéstrese que un rayo de luz, que sale de un foco de la elipse y se refleja de la tangente, pasa por el segundo foco.
4. Demuéstrese que un rayo de luz que sale del foco de una parábola y se refleja de la tangente, pasa paralelamente al eje de la parábola.
5. Demuéstrese que un rayo de luz que sale de un foco de la hipérbola y se refleja de la tangente, aparece saliente del segundo foco.

§ 97. Superficies de segundo grado

Pasemos ahora al estudio de las superficies de segundo grado, dadas en forma de las ecuaciones (95.14). Consideraremos primero la ecuación

$$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + a_0 = 0. \qquad (97.1)$$

1.1. *El número a_0 es distinto de cero; los signos de todos los números λ_1, λ_2, λ_3 son idénticos y contrarios al signo de a_0.* El cambio habitual de los coeficientes nos da

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \qquad (97.2)$$

La superficie descrita por esta ecuación se llama *elipsoide* (fig. 97.1) y la propia ecuación (97.2), ecuación *canónica* del elipsoide. De la ecuación (97.2) se deduce que los planos de coordenadas son *planos de simetría* y el origen de coordenadas, el *centro de simetría*. Los números a, b, c se denominan *semiejes* del elipsoide. Un elipsoide es una superficie limitada, encerrada dentro de un paralelepípedo $|x| \leqslant a$, $|y| \leqslant b$, $|z| \leqslant c$. La línea de intersección del elipsoide con cualquier plano representa una elipse. En efecto, tal línea de intersección es una línea de segundo grado. Por ser el elipsoide una superficie limitada, esta línea también será limitada, pero la única línea limitada de segundo orden es una elipse.

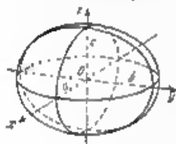


Fig. 97.1.

1.2. *El número a_0 es distinto de cero; los signos de todos los números λ_1, λ_2, λ_3, a_0 son idénticos.* La sustitución habitual de los coeficientes da

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1. \qquad (97.3)$$

No hay ningún punto del espacio cuyas coordenadas satisfagan esta ecuación. De (97.3) suele decirse como de una ecuación de un elipsoide *imaginario*.

1.3. *El número a_0 es igual a cero; los signos de todos los números λ_1, λ_2, λ_3 son idénticos.* Tenemos

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0. \qquad (97.4)$$

Esta ecuación se satisface sólo por el origen de coordenadas. Suele decirse que (97.4) es la ecuación de *una elipse degenerada*.

1.4. *El número a_4 no es igual a cero; los signos de λ_1, λ_2 coinciden y son contrarios a los de λ_3, a_4.* La sustitución habitual de los coeficientes nos da

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1. \qquad (97.5)$$

La superficie descrita por esta ecuación se llama *hiperboloide de una hoja* (fig. 97.2) y la misma ecuación, ecuación *canónica* del hiperboloide de una hoja. De la ecuación (97.5) se deduce que los planos de coordenadas son *los de simetría* y el origen de coordenadas es el *centro de simetría*. Examinemos las líneas L_h de intersección del hiperboloide de una hoja con los planos $z = h$. La ecuación de la proyección de tal línea sobre el plano Oxy se obtiene de la ecuación (97.5), si ponemos en ésta $z = h$. Es fácil ver que esta línea es una elipse

$$\frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1,$$

donde

$$a^*=a\sqrt{1+h^2/c^2},$$

$$b^*=b\sqrt{1+h^2/c^2},$$

con la particularidad de que sus dimensiones crecen ilimitadamente cuando $h\to+\infty$. Las secciones que se obtienen al cortar el hiperboloide de una hoja por los planos Oyz y Oxz representan en sí las hipérbolas.

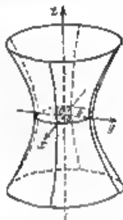


Fig. 97.2.

De este modo, el hiperboloide de una hoja representa en sí una superficie compuesta por una hoja y semejante a un tubo. Dicha superficie se extiende ilimitadamente en las direcciones positiva y negativa del eje Oz.

1.5. *El número a_4 no es igual a cero; los signos de λ_2, λ_3, a_4 coinciden y son contrarios al de λ_1.* Por analogía con (97.5), tenemos

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1. \qquad (97.6)$$

La superficie descrita por esta ecuación se denomina *hiperboloide de dos hojas* (fig. 97.3) y la propia ecuación, ecuación *canónica* del hiperboloide de dos hojas. Los planos de coordenadas son *planos de simetría* y el origen de coordenadas es *el centro de simetría*. Las líneas de intersección L_h del hiperboloide de dos hojas con los planos $z = h$ representan elipses cuyas proyecciones sobre el plano Oxy tienen por expresión

$$\frac{x^2}{a^{*2}}+\frac{y^2}{b^{*2}}=1,$$

donde

$$a^* = a\sqrt{-1+h^2/c^2}, \qquad b^* = b\sqrt{-1+h^2/c^2}.$$

De aquí se infiere que el plano secante $z = h$ empieza a cortar el hiperboloide de dos hojas sólo cuando $|h| \geqslant c$. En la capa entre los planos $z = -c$ y $z = +c$ no hay puntos de la superficie en con-



Fig. 97.3.

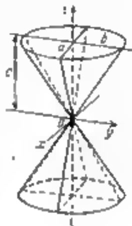


Fig. 97.4

sideración. En virtud de la simetría respecto del plano Oxy, la superficie consta de dos hojas dispuestas fuera de la capa citada. Las secciones obtenidas como resultado del corte del hiperboloide por los planos Oyz y Oxz representan en sí unas hipérbolas.

1.6. *El número a_0 es igual a cero; los signos de λ_1, λ_2 coinciden y son contrarios al de λ_3.* Tenemos

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0. \tag{97.7}$$

La superficie definida por esta ecuación se llama *cono elíptico* (fig. 97.4) y la propia ecuación, ecuación *canónica* del cono elíptico. Los planos de coordenadas sirven de *planos de simetría* y el origen de coordenadas es el *centro de simetría*. Las líneas de intersección L_h del cono elíptico con los planos $z = h$ representan en sí elipses. Si el punto $M(x_0, y_0, z_0)$ se dispone en la superficie del cono, entonces las coordenadas del punto $M_t(tx_0, ty_0, tz_0)$, para todo número t, satisfacen la ecuación (97.7). Por consiguiente, toda la recta que pasa por el punto M_0 y el origen de coordenadas se halla *íntegramente* en la superficie dada.

Pasemos ahora a considerar la segunda ecuación de (95.14). Tenemos

$$\lambda_1 x^2 + \lambda_2 y^2 + b_0 z = 0.$$

11.7. *Los números λ_1, λ_2 son de un mismo signo.* Sin limitar la generalidad, podemos considerar que b_0 tiene signo opuesto, ya que al coincidir los signos de b_0 y λ_1, λ_2 obtenemos una superficie dispuesta simétricamente respecto del plano Oxy. El cambio habitual de los coeficientes nos da

$$z = \frac{x^2}{a^2} + \frac{y^2}{b^2}. \tag{97.8}$$

La superficie descrita por esta ecuación se llama *paraboloide elíptico* (fig. 97.5) y la propia ecuación, ecuación *canónica* del paraboloide elíptico. Para esta superficie Oxz y Oyz son *los planos de simetría*, el centro de simetría no existe. El paraboloide elíptico se

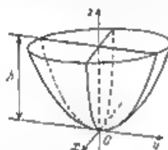


Fig. 97.5.



Fig. 97.6.

dispone en el semiespacio $z \geqslant 0$. Las líneas de intersección L_h del paraboloide elíptico con los planos $z = h$, $h > 0$, representan en sí unas elipses cuyas proyecciones sobre el plano Oxy se definen por la ecuación

$$\frac{x^2}{a^{*2}} + \frac{y^2}{b^{*2}} = 1,$$

donde $a^* = a\sqrt{h}$, $b^* = b\sqrt{h}$. De aquí se deduce que al crecer h, las elipses aumentan ilimitadamente, es decir, el paraboloide elíptico representa en sí una taza infinita.

Las secciones que se obtienen al cortar el paraboloide elíptico por los planos $y = h$ y $x = h$, representan en sí unas parábolas. Por ejemplo, el plano $x = h$ interseca la superficie a lo largo de la parábola

$$z - \frac{h^2}{a^2} = \frac{y^2}{b^2},$$

dispuesta en el plano $x = h$.

11.8. *Los números λ_1, λ_2 son de signos diferentes.* La superficie tipo para este caso se determina por la ecuación

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2}.$$

La superficie descrita por esta ecuación se llama *paraboloide hiperbólico* (fig. 97.6) y la propia ecuación, ecuación *canónica* del paraboloide hiperbólico. Los planos Oxz y Oyz son los planos de simetría, el centro de simetría no existe. Las líneas de intersección del paraboloide hiperbólico con los planos $z = h$ representan en sí, para $h > 0$, las hipérbolas

$$\frac{x^2}{a^{*2}} - \frac{y^2}{b^{*2}} = 1,$$

donde $a^* = a\sqrt{h}$, $b^* = b\sqrt{h}$, y, para $h < 0$, las hipérbolas

$$-\frac{x^2}{a^{*2}} + \frac{y^2}{b^{*2}} = 1,$$

donde $a^* = a\sqrt{-h}$, $b^* = b\sqrt{-h}$. El plano $z = 0$ corta al paraboloide hiperbólico a lo largo de dos rectas

$$y = \pm \frac{b}{a} x.$$

Todas las superficies definidas por las ecuaciones III—V de (95.14) no dependen de z. Por ello, las proyecciones de las líneas de intersección de dichas superficies con los planos $z = h$ sobre el plano Oxy tampoco dependen de h. Las superficies de tal género se llaman *cilindros*, añadiéndose la definición *elíptico*, *hiperbólico*, etc., según sea la forma de la proyección de la superficie sobre el plano Oxy.

TEOREMA 97.1 *Por todo punto del hiperboloide de una hoja y del paraboloide hiperbólico pasan dos rectas diferentes, dispuestas íntegramente en las superficies indicadas.*

DEMOSTRACIÓN Examinemos un hiperboloide de una hoja definido por su ecuación canónica

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \qquad (97.10)$$

Con cualesquiera α, β distintos de cero a la vez, un par de planos

$$\alpha\left(\frac{x}{a} + \frac{z}{c}\right) = \beta\left(1 - \frac{y}{b}\right), \qquad \beta\left(\frac{x}{a} - \frac{z}{c}\right) = \alpha\left(1 + \frac{y}{b}\right) \qquad (97.11)$$

determina cierta recta Γ. Es fácil comprobar que la recta dada Γ se dispone íntegramente en la superficie (97.10). Más aún, por todo punto de esta superficie pasa una recta perteneciente a la familia de Γ.

En efecto, consideraremos (97.11) como un sistema de dos ecuaciones

$$\alpha\left(\frac{x}{a} + \frac{z}{c}\right) - \beta\left(1 - \frac{y}{b}\right) = 0,$$

$$\alpha\left(1 + \frac{y}{b}\right) - \beta\left(\frac{x}{a} - \frac{z}{c}\right) = 0$$

respecto de α, β. El determinante del sistema es igual a cero cuando, y sólo cuando, el punto $M(x, y, z)$ se dispone en el hiperboloide (97.10). Además, el rango de la matriz del sistema es igual a uno a ciencia cierta. Por consiguiente, α y β se determinan, salvo la proporcionalidad. Pero, esto significa precisamente la unicidad de la recta Γ que pasa por todo punto del hiperboloide.

Del modo análogo nos convencemos de que por todo punto del hiperboloide pasa una única recta Γ^*, definida por los planos

$$\nu\left(\frac{x}{a}+\frac{z}{c}\right)=\lambda\left(1+\frac{y}{b}\right), \qquad \lambda\left(\frac{x}{a}-\frac{z}{c}\right)=\nu\left(1-\frac{y}{b}\right).$$

Las rectas Γ y Γ^* son distintas. Los mismos razonamientos muestran que un hiperboloide hiperbólico

$$z=\frac{x^2}{a^2}-\frac{y^2}{b^2}$$

está cubierto por dos familias diferentes de rectas Π y Π^* las cuales vienen definidas por los planos

$$\alpha z=\beta\left(\frac{x}{a}+\frac{y}{b}\right), \qquad \beta=\alpha\left(\frac{x}{a}-\frac{y}{b}\right)$$

y

$$\nu z=\lambda\left(\frac{x}{a}-\frac{y}{b}\right), \qquad \lambda=\nu\left(\frac{x}{a}+\frac{y}{b}\right).$$

Ejercicios.

1. ¿Qué representa en sí un hiperplano diametral conjugado de una dirección dada, para las superficies de segundo grado?
2. Escríbanse las ecuaciones de un plano tangente para diferentes superficies de segundo grado.
3. Investíguense las propiedades ópticas de las superficies de segundo grado.

CAPÍTULO 12 ESPACIOS BILINEALES MÉTRICOS

§ 98. Matriz y determinante de Gram

Supongamos que en un espacio lineal K_n, definido sobre el campo numérico P se ha introducido cierta forma bilineal $\varphi(x, y)$. El espacio K_n se llama *bilineal métrico*, si a cada par de vectores x, y de K_n se le ha puesto en correspondencia un número (x, y) de P denominado producto escalar, con la particularidad de que

$$(x, y) = \varphi(x, y).$$

Si una forma bilineal en el espacio complejo K_n es hermitiana, K_n se llama espacio *bilineal métrico hermitiano*. En estos casos diremos también que en el espacio lineal se ha introducido una métrica bilineal.

Puede observarse cierta analogía entre los espacios bilineales métricos y los espacios euclídeos y unitarios, considerados anteriormente. No obstante, indiquemos ahora mismo algunas diferencias sustanciales. Al comparar las definiciones del producto escalar en los espacios euclídeo y unitario con la definición de la forma bilineal, no es difícil advertir que en los espacios bilineales métricos el mencionado producto escalar puede no ser, en el caso general, simétrico y definido positivo.

El estudio de los espacios euclídeos y unitarios se reducía a la investigación de las propiedades adicionales tanto de los propios espacios como de los operadores que actúan en los espacios y surgen con relación a las formas bilineales las cuales determinan los productos escalares. El problema de estudio de los espacios bilineales métricos es el mismo. La necesidad de introducir una definición debilitada del producto escalar es debida al hecho de que no siempre las funciones bilineales, estudiadas en conjunto con los vectores del espacio y los operadores, poseen la propiedad de simetría y de definición positiva.

Muchas definiciones y hechos serán iguales tanto para los espacios bilineales métricos ordinarios como para los bilineales hermitianos. Por esta razón, siempre cuando no haya lugar para equívoca-

siones, omitiremos en lo sucesivo la palabra «hermitiano» y realizaremos los cálculos correspondientes sólo para los espacios bilineales, sobreentendiendo tácitamente que para los espacios hermitianos los cálculos se efectúan de manera análoga.

El procedimiento principal para investigar cualesquiera espacios lineales consiste en la descomposición de un vector según el sistema dado de vectores y en el estudio de dicha descomposición en función de diferentes factores. En un espacio lineal general no existe un instrumento, con ayuda del cual se podría encontrar la descomposición, pero resulta que en el espacio bilineal métrico como tal instrumento actúa el producto escalar.

Tomemos un sistema arbitrario de vectores $x_1, x_2, \ldots, x_m$ del espacio bilineal métrico K_n y un vector $x \in K_n$. Veamos qué nos da la presencia del producto escalar para investigar la posibilidad de descomponer el vector x

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m \tag{98.1}$$

según un sistema dado. Al multiplicar escalarmente la igualdad (98.1), de manera sucesiva a la derecha, por $x_1, x_2, \ldots, x_m$, obtendremos, para determinar los coeficientes desconocidos $\alpha_1, \alpha_2, \ldots, \alpha_m$ de la descomposición (98.1), un sistema de ecuaciones algebraicas lineales

$$\begin{aligned}
&\alpha_1 (x_1, x_1) + \alpha_2 (x_2, x_1) + \ldots + \alpha_m (x_m, x_1) = (x, x_1),\\
&\alpha_1 (x_1, x_2) + \alpha_2 (x_2, x_2) + \ldots + \alpha_m (x_m, x_2) = (x, x_2),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\alpha_1 (x_1, x_m) + \alpha_2 (x_2, x_m) + \ldots + \alpha_m (x_m, x_m) = (x, x_m).
\end{aligned} \tag{98.2}$$

La matriz G, que es una matriz traspuesta de este sistema, tiene la forma

$$G = \begin{pmatrix} (x_1, x_1) & (x_1, x_2) & \ldots & (x_1, x_m) \\ (x_2, x_1) & (x_2, x_2) & \ldots & (x_2, x_m) \\ \ldots & \ldots & \ldots & \ldots \\ (x_m, x_1) & (x_m, x_2) & \ldots & (x_m, x_m) \end{pmatrix} \tag{98.3}$$

y se denomina *matriz de Gram* del sistema de vectores $x_1, x_2, \ldots, x_m$. *El determinante* de esta matriz $G(x_1, x_2, \ldots, x_m)$ lleva el nombre *de Gram*. De este modo, el problema de investigación de las descomposiciones (98.1) está estrechamente vinculado con el problema de resolución de los sistemas (98.2).

Si los vectores $x_1, x_2, \ldots, x_m$ forman una base del espacio, la matriz de Gram será, para ellos, matriz de la forma bilineal principal (x, y). Las matrices de Gram para diferentes bases son congruentes y, por tanto, tienen rangos iguales. El rango de las matrices de Gram es un invariante del espacio bilineal métrico y se llama

rango de dicho espacio. La diferencia entre la dimensión y el rango del espacio se denomina *defecto* del espacio. Si el defecto es distinto de cero, el espacio bilineal métrico se llamará *degenerado*. La regularidad de la forma bilineal principal implica la regularidad de las matrices de Gram para todas las bases. En este caso el espacio bilineal métrico se denominará *regular*. Para un espacio regular el sistema (98.2), donde $x_1, \ldots, x_n$ es la base, tiene siempre una solución única, puesto que la matriz del sistema será regular y se obtiene la posibilidad de investigar los coeficientes de la descomposición (98.1) como solución del sistema (98.2).

Supongamos que para ciertos vectores x, y de un espacio bilineal métrico K_n se verifica la igualdad $(x, y) = 0$. En este caso el vector y se llama *ortogonal a la derecha* respecto del vector x, y el vector x, *ortogonal a la izquierda* respecto del vector y. En los espacios bilineales métricos hemos de distinguir las ortogonalidades a la izquierda y a la derecha, puesto que en el caso general $(x, y) \neq (y, x)$. Si, en cambio, $(x, y) = (y, x) = 0$, los vectores se llamarán simplemente *ortogonales*. Teniendo en cuenta la linealidad del producto escalar respecto de cada argumento, es fácil comprobar que la ortogonalidad del vector y a la derecha respecto de los vectores $x_1, x_2, \ldots, x_m$ origina su ortogonalidad a la derecha respecto de su cualquier combinación lineal. Lo mismo puede decirse respecto de la ortogonalidad a la izquierda. De aquí proviene, en particular, que para que un vector del espacio K_n sea ortogonal a todos los vectores del subespacio lineal L, es necesario y suficiente que sea ortogonal a los vectores de una base cualquiera del subespacio L.

LEMA 98.1. *Si la matriz de Gram del sistema de vectores $x_1, x_2, \ldots, x_m$ es degenerada, entonces existen tales vectores u, v (que son combinaciones lineales no triviales de los vectores $x_1, x_2, \ldots, x_m$) que u es ortogonal a la derecha y v, a la izquierda respecto de todos los vectores de la cápsula lineal de los vectores $x_1, x_2, \ldots, x_m$.*

DEMOSTRACIÓN. Si la matriz de Gram es degenerada, sus filas son linealmente dependientes. Por consiguiente, existen tales números $\gamma_1, \gamma_2, \ldots, \gamma_m$, no todos iguales a cero, que la combinación lineal de las filas será nula, es decir,

$$
\begin{aligned}
\gamma_1 (x_1, x_1) + \gamma_2 (x_2, x_1) + \ldots + \gamma_m (x_m, x_1) &= 0, \\
\gamma_1 (x_1, x_2) + \gamma_2 (x_2, x_2) + \ldots + \gamma_m (x_m, x_2) &= 0, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\gamma_1 (x_1, x_m) + \gamma_2 (x_2, x_m) + \ldots + \gamma_m (x_m, x_m) &= 0.
\end{aligned}
\tag{98.4}
$$

Si designamos

$$v = \sum_{j=1}^{m} \gamma_j x_j.$$

las correlaciones (98.4) significan que $(v, x_j) = 0$ para todo j. El vector v es una combinación lineal no trivial de los vectores $x_1, x_2, \ldots, x_m$, al mismo es ortogonal a la izquierda respecto a cualquiera de los vectores citados, por lo cual es ortogonal a todo vector de su cápsula lineal. El vector u se construye análogamente, pero siempre partiendo de la dependencia lineal de las columnas de la matriz de Gram.

COROLARIO. *Si la matriz de Gram para un sistema de vectores linealmente independiente es degenerada, la forma cuadrática (x, x) tiene un vector isótropo que pertenece a la cápsula lineal del sistema dado y es ortogonal a la derecha (a la izquierda) respecto de todos los vectores de esta cápsula.*

Efectivamente, en virtud de que los vectores del sistema son linealmente independientes, los vectores u, v serán no nulos; además, $(u, u) = (v, v) = 0$.

En varios casos de importancia el determinante de Gram sirve de medio muy cómodo para establecer el hecho de dependencia lineal o independencia lineal de un sistema de vectores.

LEMA 98.2. *Para todo sistema de vectores linealmente dependiente el determinante de Gram es igual a cero.*

DEMOSTRACIÓN. Sea un sistema $x_1, x_2, \ldots, x_m$ linealmente dependiente. En este caso podemos representar el vector nulo x en forma de una combinación lineal no trivial de los vectores $x_1, x_2, \ldots, x_m$. Pero, entonces, el sistema homogéneo (98.2) debe tener una solución no nula. Por consiguiente, el determinante de la matriz de este sistema, es decir, el determinante de Gram del sistema $x_1, x_2, \ldots, x_m$ será igual a cero.

TEOREMA 98.1. *Si una forma cuadrática (x, x) no tiene vectores isótropos, el determinante de Gram no es nulo, cuando, y sólo cuando, su sistema de vectores es linealmente independiente.*

DEMOSTRACIÓN. NECESIDAD. Supongamos que el determinante de Gram del sistema de vectores $x_1, x_2, \ldots, x_m$ no es igual a cero. Si suponemos que este sistema es linealmente dependiente, entonces, de acuerdo con el lema 98.2, el determinante de Gram debe ser nulo, lo que es imposible por hipótesis.

SUFICIENCIA. Supongamos que el sistema de vectores es linealmente independiente. Si el determinante de Gram es nulo, entonces, conforme al corolario del lema 98.1, debe existir un vector isótropo. Puesto que este último es imposible según la hipótesis, el determinante de Gram no es igual a cero.

COROLARIO. *Si una forma cuadrática (x, x) es estrictamente de signo constante, entonces el determinante de Gram es igual a cero cuando, y sólo cuando, el sistema de vectores es linealmente dependiente.*

COROLARIO. *Si una forma bilineal (x, x) es simétrica y la forma cuadrática (x, x) es estrictamente de signo constante, entonces para cualesquiera dos vectores x, y se verifica la desigualdad de Cauchy—Bunia-*

kovski

$$|(x, y)|^2 \leqslant (x, x)(y, y), \tag{98.5}$$

con la particularidad de que la igualdad se alcanza cuando, y sólo cuando, los vectores x, y son linealmente dependientes.

A condiciones de esta afirmación el determinante de Gram para los vectores linealmente independientes x, y será positivo, de conformidad con el criterio de Sylvester o corolario del mismo, e igual a cero, para los vectores linealmente dependientes, de acuerdo con el lema 98.2. En ambos casos la desigualdad (98.5) tiene lugar. Si, en cambio, en (98.5) se alcanza una igualdad, los vectores x, y serán linealmente dependientes, en concordancia con el corolario anterior, puesto que será nulo su determinante de Gram.

Consideraremos algunas propiedades del determinante de Gram que son sencillas, pero bastante importantes. Estas propiedades no sólo generan numerosos corolarios, sino permiten frecuentemente atribuirles una clara interpretación geométrica.

PROPIEDAD 1. *El determinante de Gram no varía, al cambiar de lugar cualesquiera dos vectores en el sistema $x_1, x_2, \ldots, x_m$.*

En efecto, si en el sistema $x_1, x_2, \ldots, x_m$ cambiamos de lugar cualesquiera dos vectores x_i y x_j, en el determinante de Gram cambiarán de lugar entre sí la i-ésima y la j-ésima columnas, como también la i-ésima y la j-ésima filas. Además, el determinante de Gram cambiará de signo dos veces y, como resultado, quedará inalterable.

PROPIEDAD 2. *El determinante de Gram no varía cuando a un vector cualquiera del sistema $x_1, x_2, \ldots, x_m$ se le adiciona cualquier combinación lineal de los demás vectores.*

Evidentemente, es suficiente considerar un caso en que varía el vector x_1, puesto que todos los casos restantes se reducen, teniendo presente la propiedad 1, a este primer caso. Supongamos que al vector x_1 se le agrega el vector $\alpha_2 x_2 + \ldots + \alpha_m x_m$. Supongamos además, que la forma bilineal (x, y) es ordinaria. Es fácil comprobar, que el nuevo determinante de Gram se obtiene del inicial, sumando a la primera fila la fila segunda multiplicada por α_2, etc. hasta la última fila multiplicada por α_m, y a la primera columna la columna segunda multiplicada por α_2, etc., hasta la última columna multiplicada por α_m. Como resultado de tal procedimiento, según se sabe, el determinante no varía. Si la forma bilineal (x, y) es hermitiana, las columnas se multiplican por $\bar{\alpha}_2, \ldots, \bar{\alpha}_m$.

PROPIEDAD 3. *Si un vector del sistema $x_1, x_2, \ldots, x_m$ se multiplica por el número α, entonces el determinante de Gram queda multiplicado por α^2, siempre que la forma bilineal (x, y) sea ordinaria, y por $|\alpha|^2$, si la forma (x, y) sea hermitiana.*

Esta vez también resulta suficiente considerar el caso de variación del vector x_1. Pero la multiplicación del vector x_1 por el número α conduce a la multiplicación de la primera fila y la primera colum-

es del determinante de Gram por el número α sólo en el caso en que la forma bilineal (x, y) sea ordinaria. En cambio, si la forma (x, y) es hermitiana, entonces la primera fila del determinante de Gram queda multiplicada por el número α, mientras que la primera columna, por el número $\bar{\alpha}$. De aquí precisamente proviene la propiedad enunciada.

PROPIEDAD 4. *Si cada uno de los vectores $x_1, x_2, \ldots, x_m$ es ortogonal a la izquierda (a la derecha) respecto de todos los vectores que le anteceden, entonces para el determinante de Gram se verifica la igualdad*

$$G(x_1, x_2, \ldots, x_m) = \prod_{i=1}^{m} (x_i, x_i). \qquad (98.6)$$

Efectivamente, la ortogonalidad a la izquierda (a la derecha) de cada uno de los vectores del sistema $x_1, x_2, \ldots, x_m$ respecto de todos los vectores antecedentes lleva a que la matriz de Gram será triangular derecha (izquierda). Pero el determinante de una matriz triangular es igual al producto de los elementos diagonales, de donde se infiere (98.6).

Son de mayor interés las propiedades de la matriz y del determinante de Gram en aquellos casos cuando la forma bilineal (x, y) interviene como simétrica real o simétrica hermitiana y es definida positiva. Desde luego, estos casos significan nada más que el espacio bilineal métrico K_n es, de hecho, euclídeo o bien, correspondientemente, unitario.

En un espacio euclídeo y unitario la matriz de Gram será, para cualquier sistema básico, una matriz definida positiva de forma cuadrática (x, x). De acuerdo con el criterio de Sylvester, todos los menores principales de la matriz de Gram serán positivos. Puesto que todo sistema de vectores linealmente independiente puede ser construido de modo que se obtenga una base, de aquí se deduce que será válido el

LEMA 98.1. *En un espacio euclídeo y unitario el determinante de Gram para cualquier sistema linealmente independiente de vectores es positivo.*

En un espacio euclídeo el determinante de Gram tiene una interpretación geométrica muy simple. De esto nos dirá el

TEOREMA 98.2. *En un espacio euclídeo el determinante de Gram $G(x_1, x_2, \ldots, x_m)$ del sistema de vectores $x_1, x_2, \ldots, x_m$ es igual al cuadrado del volumen $V^2(x_1, x_2, \ldots, x_m)$ de dicho sistema de vectores.*

DEMOSTRACIÓN. Examinemos una función real $G^{1/2}(x_1, \ldots, x_m)$ de m argumentos vectoriales $x_1, x_2, \ldots, x_m$. La función satisface las propiedades A, B de (36.3), conforme a las propiedades 2, 3 del determinante de Gram. En un espacio euclídeo cada vector de cualquier sistema ortonormalizado de vectores es ortogonal a todos los

vectores antecedentes del sistema. Por esto, en concordancia con (98.6), la función $G^{1/2}(x_1, x_2, \ldots, x_m)$ satisface también la condición C de (36.3). Pero ahora del teorema 36.1 se desprende que esta función coincide con el volumen del sistema de vectores.

COROLARIO. *Para todo sistema de vectores $x_1, x_2, \ldots, x_m$ de un espacio euclídeo se verifican las desigualdades*

$$0 \leqslant G(x_1, x_2, \ldots, x_m) \leqslant \prod_{i=1}^{m} (x_i, x_i),$$

con la particularidad de que la igualdad a la izquierda se alcanza cuando, y sólo cuando, el sistema de vectores es linealmente dependiente, mientras que la igualdad a la derecha, cuando, y sólo cuando, bien el sistema de vectores es ortogonal bien contiene el vector nulo.

La validez de esta afirmación proviene del primer corolario del teorema 98.1 y de la propiedad del volumen del sistema de vectores descrita por la desigualdad de Hadamard (38.1).

COROLARIO. *Para cualquier sistema de vectores $x_1, x_2, \ldots, x_m$ de un espacio euclídeo se verifica la desigualdad*

$$G(x_1, \ldots, x_l, x_{l+1}, \ldots, x_m) \leqslant$$
$$\leqslant G(x_1, \ldots, x_l) \cdot G(x_{l+1}, \ldots, x_m)$$

con la particularidad de que la igualdad tiene lugar cuando, y sólo cuando, o bien los conjuntos de vectores $x_1, \ldots, x_l$ y $x_{l+1}, \ldots, x_m$ son ortogonales o bien uno de los conjuntos citados representa un sistema linealmente dependiente.

La demostración se basa en un análisis muy sencillo de la fórmula (35.4). Recordemos sólo lo siguiente. Si $L_1 \subseteq L_2$, donde L_1, L_2 son unos subespacios cualesquiera, entonces $|\operatorname{ort}_{L_2} x| \leqslant |\operatorname{ort}_{L_1} x|$ para todo vector x. Con ello, la igualdad tiene lugar sólo en el caso cuando $x \perp L_2$.

Ejercicios.

1. ¿Serán equivalentes los problemas de la búsqueda de las descomposiciones (98.1) y la solución de los sistemas (98.2)?
2. ¿Qué es la solución del sistema (98.2), si el vector a no pertenece a la cápsula lineal de los vectores $x_1, \ldots, x_m$?
3. ¿Cómo se representa la matriz de Gram (98.3), si:
los vectores $x_1, \ldots, x_m$ son ortogonales dos a dos,
cada uno de los vectores $x_1, \ldots, x_m$ es ortogonal a la izquierda (a la derecha) respecto de todos los vectores antecedentes (posteriores),
cada uno de los vectores $x_1, \ldots, x_m$ es ortogonal a la izquierda (a la derecha) respecto de todos los vectores posteriores (antecedentes),
cada uno de los vectores $x_{l+1}, \ldots, x_m$ es ortogonal a la izquierda (a la derecha) respecto de cada uno de los vectores $x_1, \ldots, x_l$?
4. ¿Cómo varía la matriz de Gram cuando el sistema de vectores se somete a las transformaciones elementales?

5. Demuéstrese que si en un espacio bilineal métrico ordinario de la condición $(x, y) = 0$ se sigue siempre que $(y, x) = 0$, entonces el producto escalar está dado por cualquier forma bilineal, simétrica o antisimétrica.

6. ¿Será cierta la afirmación 5 para un espacio bilineal métrico hermítico?

7. Sea G una matriz de Gram para cierta base en un espacio bilineal métrico K_n, regular hermitiano. Demuéstrese que para el operador U con la matriz $\bar{G}^{-1}G'$ en la misma base se verifica la igualdad

$$(Ux, Ux) = (x, x),$$

cualesquiera que sean los vectores $x \in K_n$.

8. Demuéstrese que para cualquier operador lineal A que actúa en el espacio euclídeo o unitario K_n la correlación

$$k(A) = \frac{G(Ax_1, \ldots, Ax_m)}{G(x_1, \ldots, x_m)}$$

no depende de los vectores $x_1, \ldots, x_m$ y es igual al producto de los cuadrados de módulos de los valores propios del operador A.

9. Demuéstrese que para todo sistema linealmente independiente de vectores $x_1, \ldots, x_m$ de un espacio euclídeo o unitario y todo vector z se verifica la desigualdad

$$\frac{G(x_1, \ldots, x_m, z)}{G(x_1, \ldots, x_m)} \leqslant \frac{G(x_1, \ldots, x_{m-1}, z)}{G(x_1, \ldots, x_{m-1})}.$$

§ 99. Subespacios regulares

Todo subespacio lineal L de K_n puede considerarse como espacio bilineal métrico respecto del mismo producto escalar que se ha introducido en K_n. En el caso general, de la regularidad de K_n no proviene la regularidad de L, y viceversa.

TEOREMA 99.1. *Para que en el espacio K_n sean regulares todos sus subespacios, es necesario y suficiente que la forma cuadrática (x, x) no tenga vectores isótropos.*

DEMOSTRACION. NECESIDAD. Supongamos que en K_n todos los subespacios son regulares. Entonces, serán regulares también todos los subespacios unidimensionales. Pero las matrices de Gram para los vectores no nulos x coinciden con el producto escalar (x, x), el cual debe ser distinto de cero, puesto que los subespacios unidimensionales son regulares.

SUFICIENCIA. Supongamos que $(x, x) \neq 0$ para todo $x \neq 0$. Consideremos un subespacio cualquiera L y una base en él $x_1, x_2, \ldots, x_m$. De conformidad con el teorema 98.1, el determinante de Gram para este sistema es distinto de cero, es decir, el subespacio L es regular.

COROLARIO. *Para que en un espacio bilineal métrico K_n sean regulares todos sus subespacios, es necesario y suficiente que todos sus subespacios unidimensionales sean regulares.*

COROLARIO. *En cualquier espacio bilineal métrico ordinario complejo existen subespacios degenerados unidimensionales.*

Para demostrar esta afirmación es suficiente recordar que en un espacio bilineal métrico ordinario complejo toda forma cuadrática tiene vectores isótropos.

Cuando una forma cuadrática tiene vectores isótropos, en el espacio bilineal métrico existirán tanto subespacios degenerados como regulares. Si la forma bilineal (x, y) es de rango r, está claro que no puede haber subespacios regulares cuya dimensión sea superior a r. Pero, los subespacios regulares de dimensión r existen. Como ejemplo podemos indicar el subespacio tendido sobre aquellos vectores de la base canónica, para los cuales la matriz de Gram coincide con la matriz M de (92.5).

Diremos que el conjunto de vectores F de un espacio bilineal métrico K_n *es ortogonal a la derecha, a la izquierda* o simplemente *ortogonal* al conjunto de vectores G de K_n, si para cada par de vectores x, y, donde $x \in F$, $y \in G$, se cumple una relación análoga de ortogonalidad. Está claro que la totalidad de todos los vectores del espacio K_n, ortogonales a la derecha (a la izquierda) respecto de cada uno de los vectores del conjunto F, es un subespacio. Se denomina *complemento ortogonal a la derecha (a la izquierda)* del conjunto F y se designa con $F^{\perp}$ $({}^{\perp}F)$.

En los espacios euclídeo y unitario los subespacios ${}^{\perp}K_n$ y $K_n^{\perp}$ coinciden y se componen sólo de un vector nulo. En los espacios bilineales métricos estos subespacios pueden ser diferentes y no es obligatorio que se compongan sólo de un vector nulo. Los subespacios ${}^{\perp}K_n$ y $K_n^{\perp}$ se denominan subespacios *nulos* en K_n, izquierdo y derecho, respectivamente.

Cabe notar que para todo conjunto de vectores F son siempre justas las inclusiones $K_n^{\perp} \subseteq F^{\perp}$, ${}^{\perp}K_n \subseteq {}^{\perp}F$, y para cualesquiera vectores de ${}^{\perp}K_n$ o $K_n^{\perp}$ las matrices de Gram resultan nulas.

TEOREMA 99.1 *Las dimensiones de los subespacios nulos izquierdo y derecho coinciden y son iguales al defecto de la forma bilineal* (x, y).

DEMOSTRACIÓN. Elijamos en K_n una base $x_1, x_2, \ldots, x_n$. Tomemos un vector arbitrario x de $K_n^{\perp}$ y representémoslo en forma de una descomposición según la base, por analogía con (98.1). La condición de pertenencia del vector x al subespacio $K_n^{\perp}$ es equivalente a las condiciones de ortogonalidad a la derecha del vector x respecto de cada uno de los vectores de la base. Pero estas condiciones conducen a la resolución del sistema homogéneo del tipo (98.2) con el fin de hallar los coeficientes de la descomposición. Se sabe (véase § 48) que el conjunto de soluciones de dicho sistema es un subconjunto cuya dimensión es igual al defecto de la matriz de Gram o, que es lo mismo, al defecto de la forma bilineal (x, y). La demostración para el subespacio nulo izquierdo se realiza de manera análoga.

COROLARIO. *Para que el espacio K_n sea regular, es necesario y suficiente que los subespacios derecho e izquierdo se compongan sólo del vector nulo.*

En los espacios euclídeo y unitario todo subespacio es ortogonal a su complemento ortogonal y determina la descomposición de todo el espacio no sólo en una recta, sino incluso en la suma ortogonal de estos subespacios. En los espacios bilineales métricos no siempre tienen lugar los hechos análogos.

TEOREMA 99.3. *Sea L un subespacio en K_n. Para que existan las descomposiciones*

$$K_n = L \dotplus L^{\perp} = L \dotplus {}^{\perp}L, \qquad (99.1)$$

es necesario y suficiente que el espacio L sea regular.

DEMOSTRACIÓN. NECESIDAD. Supongamos que las descomposiciones (99.1) tienen lugar. Consideraremos L como un espacio bilineal métrico con el mismo producto escalar que figuraba en K_n. La intersección $L \cap L^{\perp}$ es el subespacio nulo derecho en L. Puesto que las sumas (99.1) son directas, este subespacio sólo contiene el vector nulo. De acuerdo con el corolario del teorema 99.2, esto significa que el subespacio L es regular.

SUFICIENCIA. Si el subespacio L es regular, entonces la intersección $L \cap L^{\perp}$ contendrá sólo el vector nulo y resta por señalar que todo vector $a \in K_n$ puede ser representado en la forma $a = u + v$, donde $u \in L$, $v \in L^{\perp}$. Elijamos en L una base $a_1, a_2, \ldots, a_m$. Para que exista la descomposición buscada $a = u + v$, es necesario y suficiente que en L se encuentre tal vector u que $a - u$ sea ortogonal a la derecha respecto de los vectores $a_1, a_2, \ldots, a_m$. Esta vez también obtenemos un sistema de ecuaciones algebraicas lineales con la matriz de Gram para determinar los coeficientes de la descomposición del vector u según los vectores $a_1, a_2, \ldots, a_m$. La matriz citada es regular y el sistema tiene solución, es decir, el vector u existe.

Desde luego, todo lo que hemos dicho respecto al subespacio $L^{\perp}$ es válido por completo para el subespacio ${}^{\perp}L$.

COROLARIO. *Si un subespacio regular L tiene dimensión m, entonces la dimensión de los subespacios $L^{\perp}$ y ${}^{\perp}L$ es igual a $n-m$.*

Con miras a demostrar esta afirmación resulta suficiente hacer uso de la igualdad (19.1) y recordar que la dimensión de los subespacios $L \cap L^{\perp}$ y $L \cap {}^{\perp}L$ es igual a cero.

COROLARIO. *Si un subespacio regular L tiene dimensión máxima, entonces $L^{\perp} = K_n^{\perp}$, ${}^{\perp}L = {}^{\perp}K_n$.*

En efecto, sea r el rango de la forma bilineal (x, y). Como ya se ha observado, el subespacio L será de dimensión r, mientras que los subespacios $K_n^{\perp}$ y ${}^{\perp}K$ tendrán la dimensión $n-r$. Pero los subespacios $K_n^{\perp}$ y ${}^{\perp}K_n$ tienen también esta misma dimensión $n - r$, y, además, $K_n^{\perp} \subseteq L^{\perp}$, ${}^{\perp}K_n \subseteq {}^{\perp}L$. Por eso, $K_n^{\perp} = L^{\perp}$, ${}^{\perp}K_n = {}^{\perp}L$.

En cuanto a las descomposiciones del tipo (99.1) en sumas ortogonales, cabe notar que del teorema 99.3 proviene el

COROLARIO. *Sea L un subespacio regular de dimensión máxima. Las descomposiciones (99.1) serán ortogonales cuando, y sólo cuando, los subespacios nulos izquierdo y derecho coinciden.*

Efectivamente, si las descomposiciones (99.1) son ortogonales, entonces $L^{\perp}$ es ortogonal a L no sólo a la derecha sino también a la izquierda, es decir, $L^{\perp} \subseteq {}^{\perp}L$. Por analogía, tenemos ${}^{\perp}L \subseteq L^{\perp}$. Por consiguiente, $L^{\perp} = {}^{\perp}L$. El hecho de que L es de dimensión máxima significa que $K_n^{\perp} = {}^{\perp}K_n$. En cambio, si los subespacios nulos coinciden, de aquí se infiere que $L^{\perp} = {}^{\perp}L$, es decir, los subespacios $L^{\perp}$ y ${}^{\perp}L$ son ortogonales a L tanto a la derecha como a la izquierda y las descomposiciones (98.5) son ortogonales.

Ahora podemos dar la respuesta a la pregunta sobre la relación existente entre la descomposición (98.1) y la solución del sistema (98.2). Supongamos que los vectores $x_1, \ldots, x_m$ forman la base de un subespacio regular L. De acuerdo con el teorema 99.3, tienen lugar las descomposiciones directas (99.1). Por ello, todo vector x del espacio bilineal métrico K_n puede ser representado de manera única en la forma $x = u + v$, donde $u \in L$, $v \in {}^{\perp}L$. Recordemos que el vector u se llama proyección del vector x sobre el subespacio L paralelamente al subespacio ${}^{\perp}L$. Si resolvemos el sistema de ecuaciones algebraicas lineales (98.2) y componemos el vector

$$u = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m, \qquad (99.2)$$

precisamente este vector será la proyección de x sobre el subespacio L paralelamente al subespacio ${}^{\perp}L$. Efectivamente, el vector u pertenece a L, mientras que la diferencia $x - u$, de conformidad con (98.2), es ortogonal a la izquierda respecto de los vectores $x_1, \ldots, x_m$. Por lo tanto, $x - u$ pertenece a ${}^{\perp}L$. Está claro que para proyectar el vector x sobre el subespacio L paralelamente a $L^{\perp}$, es necesario resolver el siguiente sistema

$$\begin{array}{l} \alpha_1 (x_1, x_1) + \alpha_2 (x_1, x_2) + \ldots + \alpha_m (x_1, x_m) = (x_1, x), \\ \alpha_1 (x_2, x_1) + \alpha_2 (x_2, x_2) + \ldots + \alpha_m (x_2, x_m) = (x_2, x), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \alpha_1 (x_m, x_1) + \alpha_2 (x_m, x_2) + \ldots + \alpha_m (x_m, x_m) = (x_m, x) \end{array} \qquad (99.3)$$

y después calcular la proyección buscada según (99.2). En el caso de un espacio bilineal métrico hermitiano los coeficientes α_j en (92.2) se sustituyen por $\bar{\alpha}_j$.

Ejercicios.

1. Descríbanse todos los subespacios regulares de dimensión máxima.
2. Demuéstrese que para todo conjunto L tienen lugar las inclusiones

$$L \subseteq {}^{\perp}(L^{\perp}), \qquad L \subseteq ({}^{\perp}L)^{\perp}.$$

¿En qué casos se alcanzan las igualdades en estas fórmulas?

3. Demuéstrese que si L es un subespacio regular de dimensión máxima en el espacio K_n, entonces

$$^{\perp}(L^{\perp}) = (^{\perp}L)^{\perp} = K_n.$$

4. Demuéstrese que si un producto escalar está dado mediante una forma bilineal simétrica o antisimétrica, entonces para cualquier conjunto F se verifica la igualdad $F^{\perp} = {}^{\perp}F$.

5. ¿De qué modo están ligadas entre sí la descomposición (98.1) y la solución del sistema (98.2), si la matriz de Gram del sistema $e_1, e_2, \ldots, e_m$ es degenerada?

6. ¿Puede existir en un espacio regular una base compuesta por los vectores isótropos?

7. ¿Qué puede decirse sobre un producto escalar, si las proyecciones sobre el subespacio fijado L paralelamente a ${}^{\perp}L$ y $L^{\perp}$ coinciden para todos los vectores?

8. ¿Qué puede decirse sobre un producto escalar, si las proyecciones de un vector fijado sobre todos los subespacios L paralelamente a ${}^{\perp}L$ y $L^{\perp}$ coinciden?

9. Sea L un subespacio regular del espacio bilineal métrico hermitiano K_n de rango $r < n$. Demuéstrese la equivalencia de las siguientes afirmaciones:
el subespacio ${}^{\perp}L$ es de dimensión $n - r$,
el subespacio $L^{\perp}$ es de dimensión $n - r$,
el subespacio L es de dimensión r,
los subespacios $L^{\perp}$ y $K_n^{\perp}$ coinciden,
los subespacios ${}^{\perp}L$ y ${}^{\perp}K_n$ coinciden,
el subespacio ${}^{\perp}L$ se compone de los vectores isótropos y el vector nulo,
el subespacio $L^{\perp}$ se compone de los vectores isótropos y el vector nulo,
el producto escalar en el subespacio ${}^{\perp}L$ es igual a cero,
el producto escalar en el subespacio $L^{\perp}$ es igual a cero.

10. ¿Qué forma tendrá la matriz de Gram para las bases compuestas por las bases de un subespacio regular L y un subespacio ${}^{\perp}L$ ($L^{\perp}$)?

§ 100. Ortogonalidad en las bases

En los espacios bilineales métricos las bases no son de igual paridad. Entre tales espacios hay algunos, para los cuales los sistemas (98.2) se resuelven y se utilizan con una facilidad singular. Por ejemplo, en el caso en que la parte considerable de la matriz de Gram se compone de los elementos nulos. Según el tipo que tengan las matrices de Gram, consideraremos varias clases de bases en los espacios bilineales métricos.

Las matrices más sencillas son diagonales. Las matrices diagonales de Gram aparecen cuando, y sólo cuando, las bases se componen de vectores ortogonales dos a dos. Tales bases se denominarán *ortogonales*. Un sistema de vectores que formen una base ortogonal en su cápsula lineal se llamará *sistema ortogonal*.

Las bases ortogonales pueden definirse de diferente modo. La definición mediante la ortogonalidad dos a dos no siempre es cómoda para la comprobación, sobre todo en los casos en que los vectores de la base se construyen sucesivamente, a partir del primero. Por eso, resulta a veces útil emplear la siguiente definición.

Una base $e_1, e_2, \ldots, e_n$ se llama ortogonal, si cada uno de sus vectores es ortogonal a todos los vectores antecedentes.

La matriz de Gram para los vectores que satisfacen esta definición es diagonal, por lo cual ambas definiciones son equivalentes. Generalmente, en la base pueden haber tanto vectores no isótropos, como isótropos. Los vectores de una base ortogonal siempre pueden conmutarse de modo tal que los vectores no isótropos vayan primeros y los isótropos sean últimos. La forma diagonal de la matriz de Gram en este caso queda, naturalmente, inalterable.

No todo espacio bilineal métrico o espacio bilineal métrico hermitiano tiene las bases ortogonales. Si existe aunque sea una sola base ortogonal, esto es testimonio de que la matriz de la forma bilineal (x, y) es diagonal en la base dada. Por consiguiente, la matriz de la forma bilineal (x, y) en cualquier otra base debe ser congruente de la diagonal. Por supuesto, la afirmación inversa es también cierta. Por esto

Para que en un espacio bilineal métrico o en un espacio bilineal métrico hermitiano exista una base ortogonal, es necesario y suficiente que la matriz de la forma bilineal (x, y) sea congruente de la matriz diagonal. En este caso el conjunto de todas las bases ortogonales coincide, salvo la permutación de los vectores, con el conjunto de las bases canónicas de la forma bilineal (x, y).

Ahora, apoyándonos en las investigaciones efectuadas anteriormente de las formas bilineales, podemos decir que entre los espacios bilineales métricos ordinarios tienen bases ortogonales aquellos espacios y sólo aquellos en los que la forma bilineal básica (x, y) es simétrica. Entre los espacios bilineales métricos hermitianos tienen bases ortogonales aquellos que cuentan con una forma bilineal básica (x, y) hermitiana o antihermitiana, como también con la forma bilineal (x, y) que tiene la parte real o imaginaria de signo constante de la forma cuadrática (x, x).

Indiquemos ahora mismo una distinción de principio que existe entre los espacios bilineales métricos con bases ortogonales, ordinarios y hermitianos. En un espacio bilineal métrico ordinario K_n la presencia de una base ortogonal lleva tras de sí la simetría del producto escalar (x, y) y esto último asegura, a su vez, la existencia de una base ortogonal en cualquier subespacio de K_n. En un espacio bilineal métrico hermitiano, del hecho de que en el mismo existe una base ortogonal automáticamente no se desprende, en el caso general, la existencia de una base ortogonal en cualquiera de sus subespacios. No obstante, si el producto escalar está dado mediante una forma bilineal simétrica hermitiana o antisimétrica hermitiana, este corolario sigue siendo válido.

Consideraremos una base ortogonal cualquiera $e_1, e_2, \ldots, e_n$ del espacio bilineal métrico K_n. En esta base hay tantos vectores isótropos y tantos no isótropos cuales son el defecto y el rango, respectivamente, del espacio K_n. Tomando en consideración la ley de inercia de las formas cuadráticas, concluimos que si la forma bilineal

(x, y) es simétrica real o simétrica hermítica, entonces todo base ortogonal tendrá el mismo número de vectores con valores positivos y negativos de las magnitudes (e_i, e_i). Estos números son invariantes para todas las bases ortogonales en K_n. Con arreglo a esto, hablaremos del índice positivo y negativo, como también de la signatura de los espacios con la forma simétrica (x, y). En el caso de los espacios bilineales métricos con la forma asimétrica (x, y), se tratarán sólo el rango y el defecto de los espacios.

Si el espacio K_n es regular, cada base ortogonal $e_1, e_2, \ldots, e_n$ está privada de vectores isótropos. En este caso para todo vector $x \in K_n$ es válida la descomposición

$$x = \sum_{j=1}^{n} \frac{(x, e_j)}{(e_j, e_j)} e_j. \tag{100.1}$$

Efectivamente, al multiplicar sucesiva y escalarmente la igualdad

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n \tag{100.2}$$

a la derecha por los vectores $e_1, e_2, \ldots, e_n$, obtendremos

$$\alpha_j = \frac{(x, e_j)}{(e_j, e_j)}$$

para todo j. Los vectores de la base ortogonal en un espacio regular pueden normalizarse obteniéndose la base *ortonormalizada*. Para la base ortonormalizada $e_1, e_2, \ldots, e_n$ se cumplen las correlaciones $|(e_j, e_j)| = 1$, cualquiera que sea j.

En los espacios degenerados entre los vectores de cualquier base habrá necesariamente vectores isótropos. Por esta razón, la representación (100.1) para la descomposición (100.2) de los vectores del espacio ya no será válida. No obstante, en estos espacios también las bases ortogonales resultan ser bastante útiles. Como ejemplo de su empleo probemos que es lícito el

TEOREMA 100.1. *Si en un espacio provisto de un producto escalar existe una base ortogonal, los subespacios anulos derecho e izquierdo coinciden.*

DEMOSTRACIÓN. Supongamos que en el espacio K_n de rango r existe una base ortogonal $e_1, e_2, \ldots, e_n$. Convengamos en considerar que los vectores $e_1, \ldots, e_r$ son no isótropos, mientras que $e_{r+1}, \ldots, e_n$ son isótropos. Tomemos arbitrariamente un vector $x \in K_n$ y descompongámoslo de acuerdo con (100.2). Haciendo uso de la representación (100.2) y tomando en consideración la ortogonalidad de la base e isotropía de los vectores $e_{r+1}, \ldots, e_n$, es fácil establecer que $(x, e_j) = (e_j, x) = 0$ para $r < j \leqslant n$. Por consiguiente, los vectores $e_{r+1}, \ldots, e_n$ figuran simultáneamente tanto en el subespacio anulo derecho como en el anulo izquierdo. Pero los vectores $e_{r+1}, \ldots, e_n$ son linealmente independientes, como vectores de la

base, y su número equivale a la dimensión de los subespacios nulos, por lo cual ambos subespacios nulos coinciden.

COROLARIO. *Si en un espacio bilineal métrico el producto escalar viene dado por una forma bilineal simétrica o simétrica hermitiana, los subespacios nulos derecho e izquierdo coinciden.*

COROLARIO *En toda base ortogonal los vectores isótropos, y sólo ellos, forman una base del subespacio nulo común.*

COROLARIO *Si en un espacio provisto de un producto escalar existe una base ortogonal, el espacio puede ser descompuesto en una suma ortogonal de cualquier subespacio regular de dimensión máxima y un subespacio nulo.*

El último corolario significa, de hecho, que el estudio de cualesquiera espacios degenerados con bases ortogonales se reduce al estudio por separado de los subespacios regulares con bases ortogonales y de los subespacios, donde el producto escalar es igual a cero.

El conocer la base ortogonal en un espacio permite no sólo indicar la base ortogonal en el subespacio regular de dimensión máxima, sino también obtener la descomposición explícita de la proyección ortogonal de cualquier vector sobre dicho subespacio según la base ortogonal de éste. En efecto, sea $e_1, e_2, \ldots, e_n$ una base ortogonal en K_n, sean $e_1, \ldots, e_r$ los vectores no isótropos y $e_{r+1}, \ldots, e_n$, isótropos. Designemos con L el subespacio tendido sobre los vectores $e_1, \ldots, e_r$. Está claro que L es regular, tiene la dimensión máxima, $L^{\perp} = {}^{\perp}L$ y, además,

$$K_n = L \oplus L^{\perp}.$$

Todo vector x de K_n puede ser representado unívocamente en forma de la suma $x = u + v$, donde $u \in L$, $v \in L^{\perp}$. Aquí, u se llama proyección ortogonal izquierda del vector x sobre el subespacio L, mientras que v, perpendicular izquierda a dicho subespacio. Escribamos para x la descomposición (100.2) según la base $e_1, e_2, \ldots, e_n$. La fórmula (100.1) ya no es válida. No obstante, conviene notar que los primeros r sumandos en (100.2) forman el vector u y los últimos $n-r$ sumandos, el vector v. Al multiplicar la igualdad (100.2) sucesiva y escalarmente a la derecha por $e_1, \ldots, e_r$, llegamos a que

$$u = \sum_{j=1}^{r} \frac{(x, e_j)}{(e_j, e_j)} e_j.$$

La proyección v del vector x sobre el subespacio nulo se determina de una manera más simple

$$v = x - \sum_{j=1}^{r} \frac{(x, e_j)}{(e_j, e_j)} e_j.$$

Lo único que no se puede hacer ahora es hallar la descomposición del vector v según los vectores $e_{r+1}, \ldots, e_n$, haciendo uso del producto escalar, pese a que la propia descomposición existe.

Como ya se ha dicho, las bases ortogonales existen no en todo espacio bilineal métrico o espacio bilineal métrico hermitiano. Esta circunstancia nos obliga a buscar otras clases de las bases que sean más cómodas desde el punto de vista del producto escalar prefijado en el espacio. La solución se dicta por la expresión canónica para la matriz de la forma bilineal.

La base $e_1, e_2, \ldots, e_n$ se denomina *seudoortogonal*, si cada uno de sus vectores es ortogonal a la izquierda respecto de todos los vectores antecedentes y cada uno de sus vectores isótropos es ortogonal a la izquierda respecto de todos los vectores de la base. Un sistema de vectores que forman la base seudoortogonal en su cápsula lineal se llamará *seudoortogonal*.

Hemos de notar que en la definición dada la ortogonalidad de los vectores a la izquierda respecto de todos los anteriores puede sustituirse por la ortogonalidad de los vectores a la derecha respecto de todos los vectores posteriores. Esto determina las mismas condiciones.

La matriz de Gram para los vectores de una base seudoortogonal es *trapezoidal* derecha. Si los vectores de la base se permutan de modo tal que los vectores no isótropos vayan primeros y los isótropos sean últimos, entonces la matriz de Gram no sólo queda trapezoidal derecha, sino adquiere, además, la expresión canónica (92.5). Las investigaciones realizadas anteriormente, referentes a la reducción de la matriz de una forma bilineal a la expresión canónica, nos proporcionan la respuesta completa a la pregunta sobre las condiciones de existencia de una base seudoortogonal.

La base seudoortogonal existe en cualquier espacio bilineal métrico hermitiano, como también en todo espacio bilineal métrico ordinario, a excepción de los espacios con la forma bilineal antisimétrica (x, y). El conjunto de todas las bases seudoortogonales coincide, salvo la permutación de los vectores, con el conjunto de las bases canónicas de la forma bilineal (x, y).

Toda base ortogonal es seudoortogonal. En un espacio bilineal métrico ordinario no pueden existir a la vez una base ortogonal y una seudoortogonal que no sea ortogonal. Esto se debe a que la existencia de aunque sea una sola base ortogonal lleva tras de sí la simetría de todas las matrices de Gram. La matriz trapezoidal derecha puede ser simétrica sólo en el caso, si es diagonal. En un espacio bilineal métrico hermitiano pueden existir simultáneamente una base ortogonal y una seudoortogonal que no sea ortogonal. Esto significa que la matriz compleja trapezoidal derecha puede ser congruente según Hermite a la matriz diagonal, lo que se confirma también por el ejemplo (92.8).

Si el espacio K_n es regular, toda base seudoortogonal no tiene vectores isótropos, puesto que la matriz trapezoidal derecha puede ser regular sólo en el caso cuando represente en sí una matriz trian-

gular derecha con elementos diagonales no nulos. En un espacio regular para los coeficientes α_j de la descomposición (100.2) del vector x según los vectores de la base seudoortogonal $e_1, e_2, \ldots, e_n$ obtenemos un sistema de ecuaciones algebraicas lineales con una matriz triangular izquierda. En efecto, multiplicando sucesivamente la igualdad (100.2) a la derecha por $e_1, e_2, \ldots, e_n$, encontramos que

$$\begin{aligned} \alpha_1 (e_1, e_1) \qquad\qquad\qquad\qquad\qquad &= (x, e_1) \\ \alpha_1 (e_1, e_2) + \alpha_2 (e_2, e_2) \qquad\qquad &= (x, e_2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \alpha_1 (e_1, e_n) + \alpha_2 (e_2, e_n) + \ldots + \alpha_n (e_n, e_n) &= (x, e_n). \end{aligned} \qquad (100.3)$$

De aquí determinamos sucesivamente $\alpha_1, \alpha_2, \ldots, \alpha_n$. Por supuesto en el espacio regular se puede normalizar los vectores de la base seudoortogonal y de este modo obtener la base seudoortonormalizada, para la cual $|(e_j, e_j)| = 1$, cualquiera que sea j.

Cabe notar que el proceso de resolución del sistema (100.3) da mucho más que simplemente la descomposición del vector x según la base seudoortonormalizada $e_1, e_2, \ldots, e_n$. Podemos calcular de paso, sin esfuerzos adicionales, todos los vectores

$$u_k = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_k e_k.$$

Los vectores u_k forman una sucesión de proyecciones de un mismo vector x sobre los subespacios encajados uno dentro del otro

$$L_1 \subseteq L_2 \subseteq \ldots \subseteq L_n,$$

donde L_k es la cápsula lineal de los vectores $e_1, e_2, \ldots, e_k$. Si consideramos u_k como "aproximación" a la solución x, la ortogonalidad a la izquierda del "error" $v_k = x - u_k$ respecto del subespacio L_k significa en realidad la ortogonalidad de v_k a la izquierda respecto de $u_1, u_2, \ldots, u_k$. Todas estas cuestiones las tocaremos de nuevo más adelante.

Si el espacio K_n es degenerado, entonces, en el caso general, la existencia de una base seudoortogonal no sirve de garantía para que coincidan los subespacios nulos derecho e izquierdo y, por tanto, no se puede esperar que el espacio se descomponga en la suma ortogonal de sus subespacios. Pero el saber la base seudoortogonal permite construir con eficacia la descomposición del espacio en la suma directa (99.1).

Supongamos que en el espacio K_n de rango r existe una base seudoortogonal $e_1, e_2, \ldots, e_n$. Convengamos en considerar que los vectores $e_1, \ldots, e_r$ son no isótropos, mientras que $e_{r+1}, \ldots, e_n$ son vectores isótropos. En la base seudoortogonal los vectores isótropos son ortogonales a la izquierda respecto a todos los vectores de la base y, consecuentemente, ortogonales a la izquierda a todos los vectores del espacio K_n. Pero esto significa que los vectores isótropos de la base

seudoortogonal forman una base del subespacio nulo izquierdo ${}^{\perp}K_n$. Designemos por L la cápsula lineal de los vectores $e_1, \ldots, e_r$. De acuerdo con el corolario segundo del teorema 99.3,

$$K_n = L \dotplus {}^{\perp}L = L \dotplus {}^{\perp}K_n,$$

con la particularidad de que para ambos subespacios L y ${}^{\perp}K_n$ las bases son conocidas. Para el subespacio L la base $e_1, \ldots, e_r$ está seudoortogonal.

Así pues, el estudio de todos los espacios degenerados con una base seudoortogonal se reduce al estudio conjunto de los subespacios regulares con una base seudoortogonal y los subespacios, donde el producto escalar es igual a cero.

Todo vector de K_n puede ser representado de modo único en forma de la suma $x = u + v$, donde $u \in L$, $v \in {}^{\perp}K_n$. Si para el vector x escribimos la descomposición (100.2), con miras a determinar los coeficientes α_j, obtendremos otra vez el sistema del tipo (100.3), pero ya no con una matriz triangular izquierda regular, sino con una matriz trapezoidal izquierda. No obstante, haciendo uso de este sistema, se pueden determinar los primeros coeficientes $\alpha_1, \ldots, \alpha_r$ y llegamos a que

$$u = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_r e_r,$$

es decir, la proyección del vector x sobre el subespacio L se halla por completo, si sólo se sabe la base seudoortogonal en L. Necesariamente $v = x - u$, y tampoco podemos hallar la descomposición del vector v según los vectores $e_{r+1}, \ldots, e_n$, recurriendo al producto escalar.

La base seudoortogonal es un tipo bastante general de base, puesto que existe casi en todos los espacios. Como ya sabemos, no existe sólo en los espacios bilineales métricos ordinarios con la forma antisimétrica (x, y). Para éstos últimos espacios el tipo más cómodo de la base es evidente y es, por supuesto, la base canónica de la matriz de Gram. Hablando en general, se puede introducir un tipo de la base que cubra todos los tipos de base considerados arriba y que exista en cualquier espacio dotado de un producto escalar. Sin embargo, esta introducción ofrece pocos hechos nuevos y por ahora no nos detendremos en este problema.

Además de una base con tales o cuales relaciones de ortogonalidad entre sus vectores, nos encontraremos a veces con los pares de bases análogas.

Una base $f_1, f_2, \ldots, f_n$ se llama *base dual izquierda (derecha)* para la base $e_1, e_2, \ldots, e_n$, si $(f_i, e_j) = 0$ ($(e_j, f_i) = 0$) para $i \neq j$, y, en este caso, (f_i, e_i) ((e_i, f_i)) es igual a 1 ó 0 para cualquier valor de i.

Una base $f_1, f_2, \ldots, f_n$ se llama *seudodual izquierda (derecha)* para la base $e_1, e_2, \ldots, e_n$, si $(f_i, e_j) = 0$ ($(e_j, f_i) = 0$) para todo

$j < i$ cuando $(f_i, e_j) = 1$ $((e_j, f_i) = 1)$ y para todo j cuando $(f_i, e_i) = 0$ $((e_i, f_i) = 0)$.

Es fácil ver que la matriz de la forma bilineal (x, y) en un par de bases duales es diagonal y en un par de bases seudoduales, trapezoidal derecha (izquierda). Las cuestiones de existencia y construcción de las bases duales y seudoduales están estrechamente relacionadas con las transformaciones equivalentes (91.4) de la matriz de la forma bilineal (x, y), como también con la descomposición de dicha matriz en factores. Recurriremos a la investigación detallada de las bases de tal tipo sólo cuando sea necesario. Aquí nos limitaremos sólo a una breve exposición de las cuestiones citadas.

TEOREMA 100.2. *En todo espacio regular cada base tiene las bases duales derecha e izquierda y éstas son únicas.*

DEMOSTRACIÓN. Consideremos una base $e_1, e_2, \ldots, e_n$ en el espacio regular K_n y sea G_e la matriz de la forma bilineal (x, y) en dicha base. Según (91.4), el problema de búsqueda de la base dual izquierda (derecha) para $e_1, e_2, \ldots, e_n$ es equivalente a la definición de la matriz P (Q), para la cual $P'G_e$ (G_eQ) será una matriz unidad. Entonces, P (Q) será una matriz de la transformación de coordenadas al pasar de la base $e_1, e_2, \ldots, e_n$ a una base dual. Puesto que el espacio es regular, la matriz G_e también será regular y existe la única solución: $P = G_e'^{-1}$ $(Q = G_e^{-1})$.

COROLARIO. *En todo espacio regular cualquier base tiene las bases duales izquierda y derecha.*

Efectivamente, cada base dual izquierda (derecha) es a la vez una base seudodual izquierda (derecha).

Tomando en consideración la expresión para la matriz de la forma bilineal (x, y), es fácil establecer que si en un espacio regular se pasa de una base dual izquierda (derecha) a otra base que cuente con una matriz triangular izquierda de la transformación de coordenadas cuyos elementos diagonales son unidades, entonces la base nueva será seudodual izquierda (derecha).

Ejercicios.

1. Sea simétrico un producto escalar. En el caso de las bases no ortogonales ¿son $e_1, e_2, \ldots, e_n$ invariantes del número de vectores que tienen valores nulos, positivos y negativos de las magnitudes (e_i, e_i)?

2. ¿De qué modo se puede convertir un espacio lineal complejo o real en un espacio bilineal métrico dotado de un producto escalar simétrico en el cual se han prefijado el rango y la signatura?

3. La base ortogonal en un espacio regular no tiene vectores isótropos. ¿Podrá existir en tal espacio una base de los vectores isótropos?

4. Demuéstrese que una proyección ortogonal y una perpendicular, siendo funciones de los vectores de un espacio bilineal métrico, son operadores lineales.

5. ¿Qué forma tiene la matriz de Gram para una base seudoortogonal, si los subespacios nulos derecho e izquierdo coinciden?

6. Demuéstrese que en todo espacio bilineal métrico ordinario o hermitiano existe una base en la cual la matriz de Gram es triangular celular derecha cuyas células en la diagonal son de primero y segundo órdenes.

7. ¿De qué modo se hallan los coeficientes de la descomposición de un vector según la base, para la cual se conoce alguna base dual o seudodual?

8. Demuéstrese que en un espacio regular la matriz de la transformación de coordenadas al pasar de una base, seudodual respecto de la dada, a cualquier otra base seudodual de la misma denominación es triangular izquierda.

§ 101. Operadores y formas bilineales

Si en un espacio bilineal métrico ordinario o hermitiano actúa un operador lineal, entonces, desde luego, todos los resultados obtenidos anteriormente respecto de los operadores en un espacio real o complejo siguen siendo válidos. Por eso estudiaremos aquí sólo propiedades adicionales de los operadores relacionadas con la presencia en el espacio de un producto escalar.

Uno de los objetos más importantes es el operador conjugado. En los espacios euclídeo y unitario el operador conjugado se introducía mediante un producto escalar, mas en la investigación de sus propiedades se usaba ampliamente el hecho de existencia en el espacio de una base ortonormalizada. Ahora no podemos seguir este camino, pues en un espacio bilineal métrico general puede no haber ninguna base ortogonal. Nuestras investigaciones se realizarán en un espacio bilineal métrico hermitiano. Los cambios para el espacio bilineal métrico ordinario son muy simples.

Un operador A^* (${}^*A$) que actúa en el espacio bilineal métrico hermitiano K_n se llama *conjugado derecho* (*izquierdo*) para el operador A que actúa en K_n, si para cualesquiera vectores $x, y \in K_n$ se verifica la igualdad

$$(Ax, y) = (x, A^*y) \qquad ((x, Ay) = ({}^*Ax, y)). \qquad (101.1)$$

Tomemos una base arbitraria $e_1, e_2, \ldots, e_n$ en K_n y sea G_e la matriz de Gram para dicha base. Designemos con A_e la matriz del operador A en la base $e_1, e_2, \ldots, e_n$ y mediante A_e^* y ${}^*A_e$, las matrices de los operadores A^* y ${}^*A$, siempre que existan.

TEOREMA 105.1. *Para todo operador lineal A que actúa en un espacio bilineal métrico hermitiano regular existen los únicos operadores conjugados A^* y ${}^*A$, con la particularidad de que*

$$A_e^* = \bar{G}_e^{-1} \bar{A}_e' \bar{G}_e, \qquad {}^*A_e = G_e^{-1\prime} \bar{A}_e' G_e'. \qquad (101.2)$$

DEMOSTRACIÓN. Si el operador A^* existe, entonces, con arreglo a la fórmula (101.1) y teniendo en cuenta las anotaciones matriciales del tipo (61.2), (91.7), resulta

$$(Ax, y) = (Ax)_e' G_e \bar{y}_e = x_e' (A_e' G_e) \bar{y}_e,$$

$$(x, A^*y) = x_e' G_e \overline{(A^*y)_e} = x_e' (G_e \bar{A}_e^*) \bar{y}_e.$$

Los segundos miembros de estas correlaciones deben coincidir para todos los vectores x_e, y_e, por lo cual $A'_e G_e = G_e \bar{A}^*_e$, de donde se desprende la primera igualdad de (101.2). Por analogía,

$$(x, Ay) = x'_e G_e (\overline{Ay})_e = x'_e (G_e \bar{A}_e) \bar{y}_e,$$

$$({}^*Ax, y) = ({}^*Ax)'_e G_e \bar{y}_e = x'_e ({}^*A'_e G_e) \bar{y}_e$$

y, por eso, $G_e \bar{A}_e = {}^*A'_e G_e$ y obtenemos la segunda igualdad (101.2).

Las igualdades (101.2) significan que si los operadores conjugados existen, son únicos. Ahora convengamos en considerar estas igualdades como forma para definir los operadores conjugados derecho e izquierdo. Es fácil comprobar inmediatamente que los operadores construidos de tal modo son lineales y satisfacen las correlaciones (101.1).

COROLARIO. *Si una forma bilineal hermítica (x, y) es simétrica o antisimétrica, los operadores conjugados derecho e izquierdo coinciden.*

En efecto, en estos casos las igualdades $G_e = \pm \bar{G}'_e$ se verifican, cualquiera que sea la base $e_1, e_2, \ldots, e_n$. En concordancia con (101.2), concluimos que $A^*_e = {}^*A_e$.

De este corolario se deduce que los operadores conjugados derecho e izquierdo coinciden en un espacio unitario. Se puede establecer también este hecho de otro modo. Si en un espacio unitario tomamos la base ortonormalizada $e_1, e_2, \ldots, e_n$, para ésta tiene lugar la igualdad $G_e = E$ y obtenemos las igualdades bien conocidas $A^*_e = {}^*A_e = \bar{A}'_e$.

Los operadores conjugados están ligados con el operador A por unas correlaciones determinadas. Demos a conocer algunas de ellas, por ejemplo, para el operador conjugado derecho:

$$(A + B)^* = A^* + B^*,$$

$$(\alpha A)^* = \bar{\alpha} A^*, \qquad (AB)^* = B^* A^*, \qquad (A^*)^{-1} = (A^{-1})^*. \quad (101.3)$$

Para el operador conjugado izquierdo las correlaciones son análogas. Todas las correlaciones se demuestran siguiendo el mismo esquema, empleándose las representaciones (101.2) para las matrices de los operadores conjugados. Por ello nuestro intento es sólo demostrar la validez de la última propiedad. Se tiene

$$(A^*_e)^{-1} = \bar{G}_e^{-1} (\bar{A}'_e)^{-1} (\bar{G}_e^{-1})^{-1} = \bar{G}_e^{-1} (\overline{A_e^{-1}})' \bar{G}_e = (A_e^{-1})^*.$$

Comparando las fórmulas (75.4), (101.3) se puede advertir la ausencia en (101.3) del análogo de la primera correspondencia (75.4). Ahora dicha correspondencia tiene la forma:

$$({}^*A)^* = {}^*(A^*) = A. \quad (101.4)$$

Con el objeto de demostrar que es verídico, recurrimos otra vez a las representaciones (101.2) y obtenemos

$$({}^*A_f)^* = \bar G_f^{-1} (\overline{{}^*A_e})' \bar G_f = \bar G_f^{-1} (\overline{G_e^{-1} \bar A_e' G_e})' \bar G_f = \bar G_f^{-1} \bar G_e A_e \bar G_e^{-1} \bar G_e = A_e,$$

$${}^*(A_r^*) = G_r^{-1'} (\bar A_e^*)' G_r' = G_r^{-1'} (\overline{\bar G_e^{-1} \bar A_e' \bar G_e})' G_r' = G_e^{-1'} G_e' A_e G_e^{-1'} G_r' = A_e,$$

es decir, las correlaciones (101.4) son realmente justas.

TEOREMA 101.2. *Si en un espacio bilineal métrico hermitiano regular el operador A tiene en cierta base la matriz J, entonces en la base dual derecha (izquierda) el operador A^* (${}^*A$) cuenta con la matriz J^*.*

DEMOSTRACIÓN. Supongamos que el operador A tiene una matriz J en la base $e_1, e_2, \ldots, e_n$. Examinemos la base dual derecha $f_1, f_2, \ldots, f_n$. Designemos mediante G_e, G_f y $G_{ef} = E$ las matrices de la forma bilineal (x, y) en las bases correspondientes. Si P es la matriz de la transformación de coordenadas, al pasar de la primera base a la segunda, resulta

$$G_f = G_{ef} \bar P^{-1} = \bar P^{-1}, \qquad G_e = P' G_{ef} = P',$$

y luego, al tomar en consideración (63.7), (101.2), obtenemos

$$A_f^* = \bar G_f^{-1} \bar A_f' \bar G_f = \bar G_f^{-1} (\overline{P^{-1} J P})' \bar G_f = \bar G_f^{-1} \bar G_f \bar J' \bar G_f^{-1} \bar G_f = J^*.$$

En cambio, si el operador A tiene la matriz J en la base $f_1, f_2, \ldots, f_n$, entonces, para dicha base, la base $r_1, r_2, \ldots, r_n$ será dual izquierda y ahora tenemos

$${}^*A_r = G_r^{-1'} \bar A_r' G_r' = G_r^{-1'} (\overline{P A_f P^{-1}})' G_r' = G_r^{-1'} G_r' \bar J' G_r^{-1'} G_r' = J^*.$$

El teorema demostrado es de la misma significación al investigar los operadores conjugados en espacios bilineales métricos hermitianos que el teorema 75.2 en los apartados anteriores. En particular, de este teorema se infiere que los operadores conjugados derecho e izquierdo A^* y ${}^*A$ poseen los mismos valores propios complejos conjugados respecto de los valores propios del operador A, y que los operadores conjugados derecho e izquierdo A^* y ${}^*A$ son de estructura simple, siempre que tenga la misma estructura el operador A, etc.

Además del producto escalar (x, y), en un espacio bilineal métrico hermitiano pueden definirse también otras formas bilineales hermitianas. Consideraremos, por ejemplo, las funciones del tipo (Ax, y) y (x, Ay), donde A es un operador lineal arbitrario. No es difícil convencerse de que estas funciones son unas formas bilineales hermitianas. En cualquier espacio regular K_n los diferentes operadores definen formas distintas. En efecto, si A, B son operadores diferentes, por lo menos para un solo vector x se verifica la desigualdad $Ax \neq Bx$. Supongamos que para todo $y \in K_n$ se verifica la igualdad $(Ax, y) = (Bx, y)$. De aquí se desprende que $((A - B)x, y) = 0$ para todo $y \in K_n$, es decir, $(A - B)x \in {}^{\perp}K_n$. Pero, en un espacio

regular el subespacio ${}^{\perp}K_n$ sólo se compone del vector nulo y, por eso, $Ax = Bx$.

TEOREMA 101.3. *En un espacio bilineal métrico hermitiano regular K_n cualquier forma bilineal hermitiana $\varphi(x, y)$ puede ser representada de un modo único como la expresión*

$$\varphi(x, y) = (Ax, y) = (x, By),$$

donde A, B son ciertos operadores lineales que actúan en K_n.

DEMOSTRACIÓN. Elijamos en el espacio K_n una base $e_1, e_2, \ldots, e_n$ y sea G_e la matriz de Gram en dicha base y Φ_e, la matriz de la forma $\varphi(x, y)$. Tenemos

$$\varphi(x, y) = x_e'\Phi_e\bar{y}_e = x_e'\Phi_e G_e^{-1} G_e\bar{y}_e =$$
$$= (G_e^{-1'}\Phi_e' x_e)' G_e\bar{y}_e = x_e' G_e G_e^{-1}\Phi_e\bar{y}_e = x_e' G_e(\overline{\bar{G}_e^{-1}\bar{\Phi}_e y_e}).$$

Ahora las matrices A_e, B_e de los operadores buscados se determinan mediante las igualdades

$$A_e = G_e^{-1'}\Phi_e', \qquad B_e = \bar{G}_e^{-1}\bar{\Phi}_e. \tag{101.5}$$

La unicidad de los operadores A, B se ha demostrado antes.

Un operador conjugado se define en términos del producto escalar. Por esto, si en un espacio lineal se introducen productos escalares diferentes, entonces un mismo operador lineal tendrá diferentes operadores conjugados. Supongamos que en un espacio lineal, junto con el producto escalar prefijado por la forma bilineal (x, y), se introducen, además, unos productos escalares prefijados mediante las formas (Mx, y), (x, My). Indiquemos con M, al poner este índice abajo a la izquierda (a la derecha), los operadores conjugados referentes al producto escalar (Mx, y) ((x, My)).

TEOREMA 101.4. *Para cualquier operador A y un operador regular M tienen lugar las correlaciones*

$$\begin{aligned} {}_M A' &= (MAM^{-1})', & {}^{\prime}_{M}A &= M^{-1}({}'A)M, \\ A'_M &= M^{-1}A'M, & {}'A_M &= {}'(MAM^{-1}). \end{aligned} \tag{101.6}$$

DEMOSTRACIÓN. Elijamos una base cualquiera $e_1, e_2, \ldots, e_n$ y sean G_e y M_e las matrices de la forma bilineal (x, y) y del operador M en esta base, respectivamente. De acuerdo con (101.5), la matriz de la forma bilineal (Mx, y) es igual a $M_e'G_e$. Ahora, de conformidad con (101.2) encontramos

$${}_M A_e' = \overline{(M_e'G_e)^{-1}}\,\bar{A}_e'(\overline{M_e'G_e}) = \bar{G}_e^{-1}(\bar{M}_e^{-1'}\bar{A}_e'\bar{M}_e')\bar{G}_e =$$
$$= \bar{G}_e^{-1}(\overline{M_e A_e M_e^{-1}})'\bar{G}_e = (MAM^{-1})_e'.$$

$${}^{\prime}_{M}A_e = (M_e'G_e)^{-1'}\bar{A}_e'(M_e'G_e)' = M_e^{-1'}G_e^{-1'}\bar{A}_e'G_e'M_e =$$
$$= M_e^{-1}(G_e^{-1'}\bar{A}_e'G_e')M_e = M_e^{-1}({}'A_e)M_e.$$

Las igualdades matriciales obtenidas demuestran la validez del primer grupo de igualdades operacionales (101.5). El segundo grupo se desprende evidentemente del primero, si se toman en consideración la igualdad $(x, My) = ({}^*Mx, y)$ y las correlaciones (101.2).

En un espacio bilineal métrico hermitiano se consideran diferentes tipos de los operadores. El operador A se llama *hermitiano* o *autoconjugado*, si para cualesquiera $x, y \in K_n$

$$(Ax, y) = (x, Ay),$$

y *antihermitiano* o *anticonjugado*, si

$$(Ax, y) = -(x, Ay).$$

De aquí provienen las igualdades respectivas

$$A = A^* = {}^*A, \qquad A = -A^* = -{}^*A.$$

El operador A se denomina *isométrico*, si para cualesquiera $x, y \in K_n$ se tiene

$$(Ax, Ay) = (x, y).$$

Esto nos conduce a las igualdades

$${}^*AA = A{}^*A.$$

En un espacio bilineal métrico ordinario los análogos de los operadores hermitiano y antihermitiano se llaman *simétrico* y *antisimétrico*, respectivamente. En lo que sigue nos encontraremos más de una vez con los operadores que se definen mediante la igualdad

$$A^* = \alpha E + \beta A \tag{101.7}$$

para ciertos números α, β.

No todas las propiedades de los operadores del tipo especial, ni mucho menos, se pueden transferir de un espacio unitario a un espacio bilineal métrico hermitiano, aunque tienen algo en común. Aquí pasamos por alto las investigaciones de todas estas cuestiones.

Ejercicios.

1. ¿De qué modo están ligados entre sí los polinomios característicos de los operadores A, A^*, ${}^*A$?

2. Supongamos que el subespacio L es invariante respecto del operador A. Demuéstrese que el subespacio $L^{\perp}$ ($^{\perp}L$) es invariante respecto del operador A^* (${}^*A$).

3. Demuéstrese que cualquier vector propio del operador A, correspondiente al valor propio λ, es ortogonal a la izquierda (a la derecha) a todo vector propio del operador A^* (${}^*A$) que corresponde al valor propio $\mu \neq \bar{\lambda}$.

4. Demuéstrese que cualquier vector radical del operador A, correspondiente al valor propio λ, es ortogonal a la izquierda (a la derecha) a todo vector radical del operador A^* (${}^*A$) que corresponde al valor propio $\mu \neq \bar{\lambda}$.

5. Demuéstrese que los valores propios de un operador hermitiano (antihermitiano), correspondientes a los vectores propios no isótropos, son reales (imaginarios puros).

6. Demuéstrese que los módulos de los valores propios de un operador isométrico, correspondientes a los vectores propios no isótropos, son iguales a la unidad.

7. Supongamos que en un espacio regular el producto escalar es simétrico según Hermite. Demuéstrese que si el operador A es hermitiano (antihermitiano), entonces la forma bilineal (Ax, y) es simétrica (antisimétrica) según Hermite.

8. Supongamos que en un espacio regular el producto escalar es simétrico según Hermite. Demuéstrese que si la forma bilineal (Ax, y) es simétrica (antisimétrica) según Hermite, el operador A es hermitiano (antihermitiano).

9. ¿De qué modo cambian las afirmaciones de los ejercicios 7, 8, si el producto escalar es antisimétrico según Hermite?

10. Demuéstrese que si el operador A, que satisface la condición (101.7), tiene por lo menos dos valores propios distintos, entonces $|\beta| = 1$.

§ 102. Isomorfismo bilineal métrico

Al investigar los espacios euclídeos y unitarios hemos demostrado que existe, salvo un isomorfismo, un solo espacio de cada dimensión n. Para los espacios bilineales métricos los cosas resultan ser más complejas.

Introduzcamos el concepto de isomorfismo. Diremos que los espacios ordinarios o bilineales métricos hermitianos sobre un mismo campo numérico son *isomorfos*, si son isomorfos como espacios lineales, con la particularidad de que los productos escalares de los pares de vectores correspondientes son iguales entre sí.

De esta definición se deduce que en los espacios isomorfos las matrices de Gram de los sistemas de vectores correspondientes coinciden. La afirmación recíproca es también justa. Si en los espacios bilineales métricos sobre un campo numérico común existen bases con matrices de Gram coincidentes, estos espacios son isomorfos. Efectivamente, al establecer la correspondencia entre las bases con matrices iguales de Gram, aseguramos la coincidencia de los productos escalares para cualesquiera pares de vectores de las bases y, consecuentemente, para cualesquiera pares de vectores.

TEOREMA 102.1. *Los espacios bilineales métricos ordinarios (hermitianos) sobre un mismo campo numérico son isomorfos, si y sólo si, las matrices de Gram de las bases arbitrarias de estos espacios son congruentes (congruentes según Hermite).*

DEMOSTRACIÓN. NECESIDAD. Las matrices de Gram de todas las bases de un mismo espacio son congruentes, y coinciden en las bases correspondientes de unos espacios diferentes. Por ser transitiva la relación de congruencia, las matrices de Gram para las bases arbitrarias de los espacios isomorfos serán congruentes.

SUFICIENCIA. Si las matrices de Gram para las bases arbitrarias de los espacios bilineales métricos son congruentes, existen en diferentes espacios unas bases, donde las matrices de Gram coinciden. Mas, en este caso los espacios son isomorfos.

El teorema demostrado dice que el problema de clasificación de los espacios bilineales métricos es equivalente al problema de clasificación de las formas bilineales, salvo la congruencia. Examinaremos algunas clases de los espacios bilineales métricos.

Un espacio bilineal métrico real K_n se llama *seudoeuclidiano*, si el producto escalar viene dado por una forma bilineal simétrica regular.

Para una base arbitraria de un espacio seudoeuclidiano la matriz de Gram es real simétrica y, como ya sabemos, congruente de la matriz diagonal con los elementos ± 1. Esto significa que en todo espacio seudoeuclidiano existe una base en la que el producto escalar (x, y) de los vectores x, y, cuyas coordenadas son $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots$ $\ldots, \eta_n$, se da mediante la fórmula

$$(x, y) = \xi_1\eta_1 + \ldots + \xi_k\eta_k - \xi_{k+1}\eta_{k+1} - \ldots - \xi_n\eta_n.$$

Los espacios seudoeuclidianos se determinan, salvo un isomorfismo, por sus dos características: la dimensión y la signatura, los índices positivo y negativo, etc. Entre los espacios seudoeuclidianos es de mayor interés para la física un espacio cuadridimensional de índice positivo igual a uno. Este es el así llamado *espacio de Minkowski*. Representa en sí un espacio de sucesos de la teoría especial de la relatividad.

Un espacio bilineal métrico real K_n se denomina *simpléctico*, si el producto escalar viene dado por una forma bilineal antisimétrica regular.

La matriz de Gram para cualquier espacio simpléctico es antisimétrica y, debido a esta circunstancia, es congruente de la matriz celular diagonal con células del tipo $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Por esta razón la dimensión de un espacio simpléctico es siempre par y existe, salvo un isomorfismo, un solo espacio simpléctico de dimensión prefijada. En tal espacio existe una base en la que el producto escalar de los vectores x, y con las coordenadas $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_n$ es de la forma

$$(x, y) = \xi_1\eta_2 - \xi_2\eta_1 + \ldots + \xi_{n-1}\eta_n - \xi_n\eta_{n-1}.$$

Un espacio bilineal métrico complejo K_n se llama *euclidiano complejo*, si el producto escalar viene dado por una forma bilineal simétrica regular.

Para cualquier base la matriz de Gram es simétrica compleja y congruente de la matriz unidad. Existe, salvo un isomorfismo, un solo espacio euclidiano complejo de cada dimensión. En todo espacio euclidiano complejo existe una base en la que el producto escalar de los vectores x, y es de la forma

$$(x, y) = \xi_1\eta_1 + \xi_2\eta_2 + \ldots + \xi_n\eta_n.$$

Un espacio bilineal métrico hermitiano complejo se denomina *seudounitario*, si el producto escalar viene dado por una forma bilineal simétrica hermitiana regular.

La matriz de Gram para todo espacio seudounitario es hermitiana. Es congruente según Hermite de la matriz diagonal real con los elementos ± 1. Por esta razón, existe siempre una base en la que el producto escalar de los vectores x, y tiene por expresión

$$(x, y) = \xi_1\bar{\eta}_1 + \ldots + \xi_k\bar{\eta}_k - \xi_{k+1}\bar{\eta}_{k+1} - \ldots - \xi_n\bar{\eta}_n,$$

donde $\xi_1, \ldots, \xi_n$ y $\eta_1, \ldots, \eta_n$ son las coordenadas de los vectores x, y. Esta vez también, un espacio seudounitario se determina unívocamente, salvo un isomorfismo, mediante dos características suyas: la dimensión y la signatura, los índices positivo y negativo, etc.

Ejercicios

1. Demuéstrese que en los espacios isomorfos, a las bases ortogonales (seudoortogonales, duales, seudoduales) les corresponden unas bases ortogonales (seudoortogonales, duales, seudoduales).
2. Demuéstrese que en los espacios isomorfos, a los subespacios regulares les corresponden unos subespacios regulares.
3. Demuéstrese que en los espacios isomorfos una perpendicular y una proyección pasan a ser una perpendicular y una proyección, respectivamente.
4. Demuéstrese que en los espacios isomorfos los determinantes de Gram de los sistemas correspondientes de vectores son iguales.

CAPÍTULO 13 FORMAS BILINEALES EN LOS PROCESOS DE CÁLCULO

§ 103. Procesos de ortogonalización

Uno de los conceptos más importantes relacionados con cualquier espacio bilineal métrico es el de ortogonalización. Ya nos convencimos más de una vez cuán importante es el papel que desempeñan los sistemas ortogonales de vectores y sobre todo las bases ortogonales en el estudio de los espacios euclídeos y unitarios. No es menor el papel que las bases con vectores ortogonales desempeñan en otros espacios. No obstante, la mayor parte de nuestros razonamientos ha sido asociada hasta ahora con la demostración de la existencia de tales sistemas y no con los procesos de su construcción. Cierta excepción representa solamente el método general de la transformación de matrices de las formas bilineales a la forma canónica y la construcción de las bases canónicas vinculada con la transformación citada. En vista de que los sistemas ortogonales, seudoortogonales y otros análogos son muy esenciales en la construcción de los más diversos algoritmos de cálculo, examinemos ahora un proceso general que tiene por objeto la construcción de semejantes sistemas en un espacio bilineal métrico.

Supongamos que en un espacio lineal complejo K_n viene dado, con ayuda de una forma bilineal hermítica regular, el producto escalar (x, y). Consideraremos una base $e_1, e_2, \ldots, e_n$ y trataremos de construir otra base $f_1, f_2, \ldots, f_n$ que posea las siguientes propiedades:

1) las cápsulas lineales L_k de los vectores $e_1, e_2, \ldots, e_k$ y $f_1, f_2, \ldots, f_k$ coinciden para todo $k \geq 1$.

2) la base $f_1, \ldots, f_n$ es seudoortogonal. Supongamos que $(e_1, e_1) \neq 0$ y hagamos $f_1 = e_1$. Sea ya construido el sistema de los vectores seudoortogonales $f_1, \ldots, f_k$, con la particularidad de que las cápsulas lineales de estos vectores y de los vectores $e_1, \ldots, e_k$ coinciden y $(f_i, f_i) \neq 0$ para $1 \leq i \leq k$. Buscaremos el vector f_{k+1} en la forma

$$f_{k+1} = e_{k+1} + \sum_{i=1}^{k} \alpha_{i, k+1} f_i. \qquad (103.1)$$

donde $\alpha_{1,k+1}, \ldots, \alpha_{k,k+1}$ son unos coeficientes desconocidos. Las condiciones de ortogonalidad del vector f_{k+1} a la izquierda respecto a los vectores $f_1, \ldots, f_k$ ofrecen, para determinar $\alpha_{1,k+1}, \ldots, \alpha_{k,k+1}$, un sistema de ecuaciones algebraicas lineales

$$\begin{aligned}
&\alpha_{1,k+1}(f_1, f_1) &&= -(e_{k+1}, f_1),\\
&\alpha_{1,k+1}(f_1, f_2) + \alpha_{2,k+1}(f_2, f_2) &&= -(e_{k+1}, f_2),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\alpha_{1,k+1}(f_1, f_k) + \alpha_{2,k+1}(f_2, f_k) + \ldots + \alpha_{k,k+1}(f_k, f_k) &&= -(e_{k+1}, f_k).
\end{aligned} \qquad (103.2)$$

La matriz de este sistema es triangular izquierda. Por hipótesis, sus elementos diagonales se difieren de cero, debido a lo que el sistema (103.2) tiene la única solución. Está claro que el vector f_{k+1}, construido de tal modo, junto con los vectores $f_1, \ldots, f_k$ forman un sistema seudoortogonal y su cápsula lineal coincide con la de los vectores $e_1, \ldots, e_{k+1}$. El sistema de vectores $f_1, \ldots, f_{k+1}$ es linealmente independiente, puesto que linealmente independiente es el sistema $f_1, \ldots, f_k, e_{k+1}$.

Seguimos adelante el proceso. Si resulta que las magnitudes (f_i, f_i) son distintas de cero para i cualquiera, entonces el sistema obtenido de vectores $f_1, \ldots, f_n$ será precisamente la base seudoortogonal buscada. Desde luego, podemos ahora normalizar los vectores $f_1, \ldots, f_n$ y obtener una base ortonormalizada.

De la fórmula (103.1) se desprende un corolario de importancia. Escribamos la igualdad (103.1) bajo la forma

$$e_{k+1} = \left(-\sum_{i=1}^{k} \alpha_{i,k+1} f_i\right) + f_{k+1}.$$

El vector entre paréntesis pertenece a L_k, el vector f_{k+1} pertenece a $^{\perp}L_k$ por construcción, por lo cual la solución de los sistemas (103.2) nos da, en realidad, la descomposición de cada vector e_{k+1} en su proyección y la perpendicular izquierda respecto del subespacio L_k.

El proceso descrito se simplifica de modo considerable, si el producto escalar viene dado por una forma bilineal simétrica hermítica. En tal caso las condiciones $(f_j, f_i) = 0$ para $j < i$ llevan consigo el cumplimiento de las condiciones $(f_i, f_j) = 0$ para $j \neq i$. Por ello, el sistema (103.2) se convierte en un sistema con una matriz diagonal y se tiene

$$\alpha_{i,k+1} = -\frac{(e_{k+1}, f_i)}{(f_i, f_i)}$$

para todo i. La base construida $f_1, f_2, \ldots, f_n$ será no sólo seudoortogonal, sino también ortogonal.

La única causa por la cual puede obstaculizarse la construcción de la base seudoortogonal $f_1, \ldots, f_n$, a partir de la base $e_1, \ldots, e_n$, consiste en la anulación de uno de los productos escalares (f_i, f_i), $i < n$. Tal situación se llamará degenerada. La situación degenerada no viene a ciencia cierta, si la forma cuadrática (x, x) no tiene vectores isótropos, por ejemplo, si esta última es estrictamente de signo constante. Efectivamente, en este caso la igualdad $(f_i, f_i) = 0$ es posible sólo cuando $f_i = 0$. Mas, $f_i \neq 0$ para i cualquiera, puesto que los vectores $f_1, \ldots, f_i$ son linealmente independientes. Por consiguiente, ahora el proceso es realizable, cualquiera que sea la elección de la base $e_1, \ldots, e_n$.

Existen muchos problemas en los cuales no hace falta conservar los lazos de la nueva base $f_1, \ldots, f_n$ con la base inicial $e_1, \ldots, e_n$, puesto que se necesita sólo construir en el espacio una base seudoortogonal. En este caso, cada vez que aparezca la igualdad $(f_i, f_i) = 0$, se debe sustituir el vector e_i por otro y calcular de nuevo el vector f_i, repitiendo este procedimiento hasta que se cumpla la condición $(f_i, f_i) \neq 0$. Los vectores $f_1, \ldots, f_{i-1}$ quedan invariables.

Un vector necesario para sustituir e_i siempre existe. Supongamos que la igualdad $(f_i, f_i) = 0$ se verifica para cualquier vector e_i. Puesto que el vector f_i es ortogonal a la izquierda respecto a los vectores $e_1, \ldots, e_{i-1}$, esto significa que el subespacio ${}^{\perp}L_{i-1}$ está compuesto sólo por los vectores isótropos y el vector nulo. Mas, el subespacio L_{i-1} es regular, por lo cual ${}^{\perp}L_{i-1} = {}^{\perp}K_n$. La última igualdad no puede tener lugar para $i - 1 < n$, puesto que, por ser K_n regular, el subespacio ${}^{\perp}K_n$ sólo consta del vector nulo.

Al proceder del mismo modo, podemos construir también una base que sea seudodual para la dada. Supongamos otra vez que $e_1, e_2, \ldots, e_n$ es la base dada y es necesario construir una base que sea seudodual para la base dada, por ejemplo, la izquierda. Tomemos una base más: $q_1, q_2, \ldots, q_n$. Sea $(q_1, e_1) \neq 0$ y pongamos $t_1 = q_1$. Admitamos que ya se ha construido un sistema de vectores $t_1, \ldots, t_k$ tales que su cápsula lineal coincida con la cápsula lineal de los vectores $q_1, \ldots, q_k$ y se cumplan las condiciones $(t_i, e_i) \neq 0$ para $1 \leqslant i \leqslant k$ y $(t_i, e_j) = 0$ para $j < i$. Buscaremos el vector t_{k+1} en forma

$$t_{k+1} = q_{k+1} + \sum_{i=1}^{k} \beta_{i,\,k+1} t_i, \qquad (103.3)$$

donde $\beta_{1,\,k+1}, \ldots, \beta_{k,\,k+1}$ son unos coeficientes desconocidos. La condición de ortogonalidad del vector t_{k+1} a la izquierda respecto de los vectores $e_1, \ldots, e_k$ nos ofrece de nuevo, para determinar $\beta_{1,\,k+1}, \ldots, \beta_{k,\,k+1}$, un sistema de ecuaciones algebraicas lineales provisto

de una matriz triangular izquierda:

$$
\begin{aligned}
&\beta_{1,k+1}(t_1, e_1) &&= -(q_{k+1}, e_1)\\
&\beta_{1,k+1}(t_1, e_2)+\beta_{2,k+1}(t_2, e_2) &&= -(q_{k+1}, e_2)\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\beta_{1,k+1}(t_1, e_k)+\beta_{2,k+1}(t_2, e_k)+\dots+\beta_{k,k+1}(t_k, e_k) &&= -(q_{k+1}, e_k).
\end{aligned}
\tag{103.4}
$$

De acuerdo con la suposición respecto a los elementos diagonales, el sistema tiene la solución única. Si, al continuar el proceso, resulta que las magnitudes (t_i, e_i) son distintas de cero para i cualquiera, entonces el sistema obtenido de vectores será, siendo normalizado adecuadamente, la base seudodual para $e_1, e_2, \dots, e_n$. Cabe notar que esta vez el proceso no se simplifica, si el producto escalar viene dado mediante una forma bilineal simétrica. El empleo de la base auxiliar $q_1, q_2, \dots, q_n$ permite evitar la degeneración del proceso, sustituyendo en el momento apropiado uno de los vectores q_i y repitiendo los cálculos del vector t_i. En el caso dado tampoco cambian los vectores $t_1, \dots, t_{i-1}$.

Todos los procesos descritos y los procesos análogos se llamarán, con mayor frecuencia independientemente de su contenido concreto, *procesos de ortogonalización*. No obstante, a veces nos veremos obligados a construir, en un mismo espacio bilineal métrico K_n, unas sucesiones de vectores, ortogonales o seudoortogonales con relación a las diferentes formas bilineales. Se considerarán sólo las formas del tipo (Rx, y), donde R es un operador lineal en K_n. Para poder distinguir diferentes sucesiones una de la otra, diremos que se trata de la R-ortogonalización, la R-seudoortogonalización, etc.

Muchas propiedades y peculiaridades de los procesos de ortogonalización pueden establecerse considerando sus notaciones matriciales. Supongamos que un producto escalar en K_n viene dado por la forma bilineal hermitiana (x, y). La seudoortogonalidad de la base $f_1, f_2, \dots, f_n$ significa que $(f_i, f_j) = 0$ para $j < i$, es decir, la matriz de Gram G_f de la forma bilineal (x, y) en la base $f_1, f_2, \dots, f_n$ será triangular derecha. Conforme al proceso de construcción de la nueva base, las cápsulas lineales de los vectores $f_1, \dots, f_k$ y $e_1, \dots, e_k$ coinciden. Por consiguiente, al tomar en consideración (103.1), concluimos que

$$
\begin{aligned}
e_1 &= f_1,\\
e_2 &= -\alpha_{1,2}f_1 + f_2\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
e_n &= -\alpha_{1,n}f_1 - \alpha_{2,n}f_2 - \dots - \alpha_{n-1,n}f_{n-1} + f_n,
\end{aligned}
\tag{103.5}
$$

donde α_{ij} son precisamente aquellos coeficientes que se calculan partiendo de los sistemas (103.2). Por ello, la matriz A de la trans-

formación de coordenadas al pasar de la nueva base $f_1, f_2, \ldots, f_n$ a $e_1, e_2, \ldots, e_n$ es triangular derecha cuyos elementos diagonales son iguales a la unidad. Dado que la matriz de la transformación de coordenadas al pasar de la base antigua a la nueva coincide con A^{-1}, se tiene

$$G_f = A^{-1'} G_e \bar{A}^{-1}.$$

De aquí proviene la igualdad

$$G_e = A' G_f \bar{A}. \qquad (103.6)$$

Es fácil comprobar que la matriz $G_f \bar{A}$ es triangular derecha y sus elementos diagonales coinciden con los elementos diagonales de la matriz G_f.

Designemos con E_q (F_q) una matriz cuyas columnas están representadas por las coordenadas de los vectores $e_1, \ldots, e_n$ ($f_1, \ldots, f_n$) en la base $q_1, \ldots, q_n$. Las correlaciones (103.5) muestran que

$$E_q = F_q A, \qquad (103.7)$$

y, por supuesto, además,

$$G_f = F_q' G_q \bar{F}_q. \qquad (103.8)$$

Así pues, el proceso considerado de construcción de una base seudoortogonal resulta estrechamente relacionado con la descomposición de la matriz de Gram en factores triangulares y con la descomposición (103.7) en factores de la matriz de las coordenadas.

TEOREMA 103.1. *Para que el proceso* (103.1), (103.2) *de construcción de la base seudoortogonal $f_1, f_2, \ldots, f_n$ partiendo de la base $e_1, e_2, \ldots, e_n$, sea realizable en un espacio bilineal métrico regular K_n, es necesario y suficiente que la matriz de Gram del sistema $e_1, e_2, \ldots, e_n$ tenga menores principales no nulos.*

DEMOSTRACIÓN. NECESIDAD. Supongamos que el proceso es realizable, es decir, tiene lugar la correlación (103.6). La matriz G_f es regular, puesto que representa la matriz de Gram de la forma bilineal regular (x, y) para la base. Por esta razón todos sus elementos diagonales son distintos de cero. Aplicando la fórmula de Binet-Cauchy, obtenemos que para todo r

$$G_e\begin{pmatrix}1 & 2 & \ldots & r\\ 1 & 2 & \ldots & r\end{pmatrix} = A'G_f\bar{A}\begin{pmatrix}1 & 2 & \ldots & r\\ 1 & 2 & \ldots & r\end{pmatrix} = G_f\begin{pmatrix}1 & 2 & \ldots & r\\ 1 & 2 & \ldots & r\end{pmatrix} \neq 0.$$

SUFICIENCIA. Supongamos que los menores principales de la matriz de Gram G_e son distintos de cero. Por lo tanto, en concordancia con (93.1), existe una descomposición $G_e = L_e D_e U_e$, donde L_e es una matriz triangular izquierda con elementos diagonales unidad, D_e es una matriz diagonal con elementos no nulos, U_e es una matriz triangular derecha con elementos diagonales unidad. Es fácil ver

que la matriz

$$G_f = \overline{U}_e^{-1\prime} G_e U_e^{-1} = \overline{U}_e^{-1\prime} L_e D_e$$

es triangular izquierda cuyos elementos diagonales coinciden con los elementos diagonales de la matriz D_e. Ahora, si a título de la matriz A tomamos la matriz triangular derecha $\overline{U}_e$, con elementos diagonales unidad, entonces para la base $f_1, f_2, \ldots, f_n$ se cumplirán las correlaciones (103.5). Precisamente esta base será construida conforme al proceso (103.1), (103.2), de lo que es fácil convencerse por comprobación inmediata.

Si el producto escalar está dado por una forma bilineal hermítica simétrica, la matriz G_e será hermítica, como lo será también la matriz G_f. Pero de aquí se deduce que la matriz G_f será diagonal. Este hecho ya se ha notado. Comparando (93.5), (103.6), concluimos que el proceso de ortogonalización en el caso dado coincide completamente con el de la obtención de la descomposición (93.5).

Si el espacio K_n es unitario, el proceso de ortogonalización determina no sólo la descomposición de la matriz de Gram en factores triangulares, sino también la descomposición de la matriz de las coordenadas en un producto de los factores unitario y triangular derecho. En efecto, elijamos una base ortonormalizada $q_1, q_2, \ldots, q_n$ e indiquemos con D_q una matriz diagonal formada de las longitudes de las columnas de la matriz F_q de (103.7). Ahora tenemos $E_q = (F_q D_q^{-1})(D_q A)$. La matriz $D_q A$ es triangular derecha. Pero las matrices G_q y $G_f (D_q^{-1})^2$ son matrices unidad. Según (103.8), en el caso dado $(F_q D_q^{-1})' \overline{(F_q D_q^{-1})} = E$, es decir, la matriz $F_q D_q^{-1}$ es unitaria.

El hecho de que la base $t_1, t_2, \ldots, t_n$ es seudodual izquierda para la base $e_1, e_2, \ldots, e_n$ testimonia que para la forma bilineal (x, y), que determina el producto escalar en K_n, se cumplen las condiciones $(t_i, e_j) = 0$ para $j < i$ y $(t_i, e_i) = 1$ para cualquier i. En otras palabras, esto significa que para el par de bases $e_1, e_2, \ldots, e_n$ y $t_1, t_2, \ldots, t_n$ la matriz G_{te} de la forma bilineal (x, y) es triangular derecha con elementos diagonales unidad. De aquí concluimos que la matriz Q^{-1} de la transformación de coordenadas, al pasar de la base inicial $q_1, q_2, \ldots, q_n$ a la base $t_1, t_2, \ldots, t_n$, es triangular derecha. Sin embargo, en este caso los elementos diagonales no serán iguales a unidades, puesto que los vectores $t_1, t_2, \ldots, t_n$ se han sometido a la normalización. Se tiene

$$G_{te} = Q^{-1\prime} G_{qe},$$

y luego

$$G_{qe} = Q' G_{te}.$$

El proceso de construcción de la base, seudodual respecto a la dada, resulta también estrechamente ligado con la descomposición (93.1) de una matriz en factores triangulares.

TEOREMA 103.3. *Para que sea realizable el proceso de construcción* (103.3), (103.4) *de la base $t_1, t_2, \ldots, t_n$ izquierda seudodual para $e_1, e_2, \ldots, e_n$, a partir de la base $q_1, q_2, \ldots, q_n$, es necesario y suficiente que la matriz G_{qe} de la forma bilineal* (x, y) *tenga en las bases $q_1, q_2, \ldots, q_n$ y $e_1, e_2, \ldots, e_n$ los menores principales no nulos.*

La demostración de este teorema se omite aquí, puesto que es casi la repetición textual de la demostración del teorema anterior.

Subrayemos, como conclusión, que los procesos considerados de ortogonalización se extienden completamente al caso de los espacios bilineales métricos ordinarios. Cambian sólo algunos detalles relacionados con la conjugación compleja. Además, en el caso dado resulta más difícil eliminar situaciones degeneradas.

Ejercicios.

1. ¿Cuál es la interpretación geométrica del proceso de ortogonalización?
2. Demuéstrese que si el proceso de ortogonalización se aplica a un sistema linealmente dependiente $e_1, e_2, \ldots, e_n$, entonces $f_k = 0$ para cierto $k \le n$.
3. Supongamos que la forma cuadrática (x, x) está privada de vectores isótropos. ¿Cómo se determina la base del sistema dado de vectores con ayuda del proceso de ortogonalización?
4. Demuéstrese que si un proceso de ortogonalización se realiza en un espacio euclídeo o en un espacio unitario, la desigualdad $|f_k| \le |e_k|$ se cumple para k cualquiera, con la particularidad de que la igualdad se logra cuando, y sólo cuando, el vector e_k es ortogonal a los vectores $e_1, \ldots, e_{k-1}$.
5. Supongamos que las coordenadas de los vectores $e_1, \ldots, e_n$ en cierta base ortonormalizada de un espacio euclídeo o unitario forman una matriz triangular. ¿Cómo varía la matriz de las coordenadas después de realizarse el proceso de ortogonalización?
6. ¿Se podrá construir una base dual con ayuda de un proceso de ortogonalización?
7. ¿Cómo se debe aplicar el proceso de ortogonalización para obtener una base seudodual derecha?
8. Demuéstrese que la descomposición de una matriz regular compleja en el producto de una matriz unitaria y una triangular derecha es única, si se exige que los elementos diagonales de la matriz triangular sean positivos.
9. ¿Cómo se emplea el proceso de ortogonalización para la resolución de los sistemas de ecuaciones algebraicas lineales?
10. Sea K_n un espacio degenerado. ¿Cómo se construirá la base seudoortogonal de un subespacio regular de dimensión máxima, empleando para ello el proceso de ortogonalización?
11. ¿Se simplificará la construcción de los sistemas seudoortogonales de vectores en un espacio degenerado, si la forma cuadrática (x, x) es de signo constante?

§ 104. Ortogonalización de una sucesión de potencias

En los procesos de ortogonalización la matriz de la transformación de coordenadas, al pasar de una base antigua a una nueva, es siempre triangular. Sin embargo, si la base inicial se elige de un modo especial, para la matriz de la trans-

formación de coordenadas pueden obtenerse unas representaciones mucho más simples y, consecuentemente, serán más sencillos son los procesos de ortogonalización.

Supongamos que en un espacio bilineal métrico K_n, hermitiano regular, se ha dado un operador A. Tomemos un vector no nulo x y consideramos la sucesión de vectores

$$x,\ Ax,\ A^2x,\ \ldots,\ A^{k-1}x. \tag{104.1}$$

Tales sucesiones se llamarán *sucesiones de potencias generadas por el vector* x.

En toda sucesión de potencias cierto número de los primeros vectores es linealmente independiente. Supongamos que k es el mayor de estos números. Esto significa que existen tales números $\alpha_0, \alpha_1, \ldots, \alpha_k$, siendo $\alpha_k \neq 0$, que

$$\alpha_0 x + \alpha_1 Ax + \ldots + \alpha_k A^k x = 0. \tag{104.2}$$

Designemos con $\varphi(\lambda) = \alpha_k \lambda^k + \ldots + \alpha_1 \lambda + \alpha_0$ el polinomio de grado k. Por lo visto, la igualdad (104.2) es equivalente a la siguiente

$$\varphi(A)\,x = 0. \tag{104.3}$$

Existen muchos polinomios, para los cuales se verifican las correlaciones del tipo (104.3). A tales polinomios pertenecen, por ejemplo, el polinomio característico del operador A. Pero entre ellos existe, a ciencia cierta, un polinomio de grado inferior. Se denomina polinomio *mínimo* que anula al vector x. Está claro que su grado es igual al número máximo de los primeros vectores de la sucesión de potencias (104.1) que forman un sistema linealmente independiente o, lo que es igual, es inferior en una unidad al número mínimo de los primeros vectores que forman un sistema linealmente dependiente.

El grado del polinomio mínimo resulta ser íntimamente relacionado con la descomposición del vector x según la base radical del operador A, las alturas de los vectores radicales y el número de los valores propios distintos dos a dos. A saber, es verídico el

Lema 104.1. *El grado del polinomio mínimo que anula el vector x es igual a la suma de las alturas máximas de los vectores radicales del operador A que figuran en la descomposición del vector x según la base radical y que corresponden a los valores propios distintos dos a dos.*

DEMOSTRACIÓN. Representemos el vector x en forma de la suma

$$x = u_1 + u_2 + \ldots + u_t, \tag{104.4}$$

donde $u_1, \ldots, u_t$ pertenecen a los subespacios cíclicos diferentes del operador A. Puesto que los subespacios cíclicos diferentes no tienen vectores comunes, salvo el nulo, entonces para que se cumpla la igualdad (104.3), es necesario y suficiente el cumplimiento de la igualdad $\varphi(A)\,u_i = 0$ para i cualquiera. Si u_i es un vector radical de altura m_i y corresponde al valor propio λ_i, entonces la igualdad $\varphi(A)\,u_i = 0$ tendrá lugar cuando, y sólo cuando el polinomio

$\varphi(\lambda)$ se divide por $(\lambda - \lambda_1)^r$, donde $r \geqslant m_1$. En este caso $\varphi(A)\, u_j = 0$ no sólo cuando $j = 1$, sino para todos los j, para los cuales los vectores u_j corresponden a los valores propios, coincidentes con λ_1, y tienen alturas no superiores a r. Sean $\lambda_{i_1}, \ldots, \lambda_{i_p}$ unos valores propios distintos dos a dos que corresponden a los vectores $u_1, \ldots, u_l$ de (104.4) y sean $m_{i_1}, \ldots, m_{i_p}$ las alturas máximas de los vectores radicales $u_1, \ldots, u_s$ que corresponden a los valores propios y coinciden con $\lambda_{i_1}, \ldots, \lambda_{i_p}$. Entonces

$$\varphi(\lambda) = (\lambda - \lambda_{i_1})^{m_{i_1}} \ldots (\lambda - \lambda_{i_p})^{m_{i_p}}$$

será el polinomio mínimo que anula el vector x. El lema queda demostrado.

Supongamos que los vectores $e_i = A^{i-1}x$, para $1 \leqslant i \leqslant k$, son linealmente independientes. Apliquemos a este sistema el proceso, descrito anteriormente, para obtener un sistema seudoortogonal de vectores f_i, considerando, por supuesto, que el propio proceso es realizable. Si el operador A de ningún modo está ligado con el producto escalar, introducido en el espacio K_n, es difícil esperar que el proceso se simplifique. Sin embargo, la situación cambia bruscamente, si el operador A satisface la correlación (101.7), siendo, por ejemplo, autoconjugado en un espacio unitario.

TEOREMA 104.1. *Si el operador A satisface la correlación (101.7) y si los vectores $e_i = A^{i-1}x$ son linealmente independientes para $1 \leqslant i \leqslant k$, mientras que los vectores $f_1, \ldots, f_k$ se han obtenido de los vectores $e_1, \ldots, e_k$ con ayuda del proceso de seudoortogonalización, entonces tienen lugar las siguientes correlaciones*

$$\begin{aligned} f_1 &= x, \\ f_2 &= Af_1 - \alpha_1 f_1, \\ f_{i+1} &= Af_i - \alpha_i f_i - \beta_{i-1} f_{i-1}, \quad i > 1, \end{aligned} \tag{104.5}$$

donde

$$\begin{gathered} \alpha_1 = \frac{(Af_1, f_1)}{(f_1, f_1)}, \qquad \beta_{i-1} = \frac{(Af_i, f_{i-1})}{(f_{i-1}, f_{i-1})}, \quad i > 1, \\ \alpha_i = \frac{(f_{i-1}, f_{i-1})(Af_i, f_i) - (Af_i, f_{i-1})(f_{i-1}, f_i)}{(f_{i-1}, f_{i-1})(f_i, f_i)}, \\ i > 1. \end{gathered} \tag{104.6}$$

Para concretar, la demostración se realizará en un espacio bilineal métrico hermitiano. Tomando en consideración la forma de los vectores e_i y rigiéndonos por las fórmulas (103.5), concluimos que

$$f_i = A^{i-1}x + \sum_{j=0}^{i-2} \gamma_{j,i} A^j x$$

para ciertos números γ_{ji}. De aquí se desprende que el vector $f_{i+1} - Af_i$ pertenece a la cápsula lineal de los vectores $x, Ax, \ldots$

$\ldots, A^{i-1}x$ o, lo que es igual, de los vectores $f_1, f_2, \ldots, f_i$. Por otra,

$$f_{i+1} = Af_i + \sum_{j=1}^{i} \xi_{j,\,i+1} f_j$$

para ciertos números $\xi_{j,\,i+1}$. Las condiciones de ortogonalidad del vector f_{i+1} a la izquierda respecto de los vectores $f_1, f_2, \ldots, f_i$ proporcionan el siguiente sistema de ecuaciones algebraicas lineales para determinar los coeficientes $\xi_{j,\,i+1}$

$$\begin{aligned}
&\xi_{1,\,i+1}(f_1, f_1) &&= -(Af_i, f_1),\\
&\xi_{1,\,i+1}(f_1, f_2) + \xi_{2,\,i+1}(f_2, f_2) &&= -(Af_i, f_2),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots &&\\
&\xi_{1,\,i+1}(f_1, f_{i-1}) + \ldots + \xi_{i-1,\,i+1}(f_{i-1}, f_{i-1}) &&= -(Af_i, f_{i-1}),\\
&\xi_{1,\,i+1}(f_1, f_i) + \ldots + \xi_{i-1,\,i+1}(f_{i-1}, f_i) + \xi_{i,\,i+1}(f_i, f_i) &&= -(Af_i, f_i).
\end{aligned} \tag{104.7}$$

Por hipótesis del teorema, el operador A satisface la condición (101.7). Por eso, tomando en consideración la semiortogonalidad del sistema de vectores f_j, tenemos para $l < i-1$

$$(Af_i, f_l) = (f_i, A^*f_l) = (f_i, (\alpha E + \beta A) f_l) = \bar{\alpha}(f_i, f_l) + \bar{\beta}(f_i, Af_l) =$$
$$= \bar{\beta}\Big(f_i, f_{l+1} - \sum_{j=1}^{l} \xi_{j,\,l+1} f_j\Big) = \bar{\beta}\Big((f_i, f_{l+1}) - \sum_{j=1}^{l} \xi_{j,\,l+1}(f_i, f_j)\Big) = 0.$$

Entre los segundos miembros del sistema (104.7) solamente los dos últimos son distintos de cero. Por consiguiente, sólo $\xi_{i-1,\,i+1}$, $\xi_{i,\,i+1}$ pueden ser distintos de cero, lo que demuestra la validez de las correlaciones (104.5). El valor del coeficiente α_1 se halla, partiendo de las condiciones de ortogonalidad del vector f_2 a la izquierda respecto de f_1, mientras que los valores para los coeficientes α_i, β_{i-1} se obtienen de la condición de ortogonalidad del vector f_{i+1} a la izquierda respecto de f_i, f_{i-1}.

Así pues, el proceso de ortogonalización para una sucesión de potencias resulta realmente mucho más simple que para una sucesión de forma general. Si en cierto paso resulta que $(f_i, f_i) = 0$, pero $f_i \neq 0$, entonces la degeneración del proceso puede ser evitada por elección adecuada del vector nuevo x.

Supongamos que en una sucesión de potencias se hallan n vectores linealmente independientes. Aplicando el proceso de ortogonalización, se puede construir una base $f_1, f_2, \ldots, f_n$ del espacio K_n. Esta base posee la singularidad de que en ella la matriz A_f del operador A

tiene forma *tridiagonal*

$$
A_f = \begin{Vmatrix}
\alpha_1 & \beta_1 & & & & 0 \\
1 & \alpha_2 & \beta_2 & & & \\
 & 1 & \alpha_3 & \beta_3 & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & & & 1 & \alpha_{n-1} & \beta_{n-1} \\
 & & & & 1 & \alpha_n
\end{Vmatrix}. \qquad (104.8)
$$

En efecto, como columnas de la matriz del operador actúan las coordenadas de los vectores $Af_1, Af_2, \ldots, Af_n$ respecto de la base $f_1, f_2, \ldots$ $\ldots, f_n$. Pero, de acuerdo con (104.5),

$$
\begin{aligned}
Af_1 &= \alpha_1 f_1 + f_2, \\
Af_2 &= \beta_1 f_1 + \alpha_2 f_2 + f_3, \\
&\cdots\cdots\cdots\cdots\cdots \\
Af_{n-1} &= \beta_{n-2} f_{n-2} + \alpha_{n-1} f_{n-1} + f_n, \\
Af_n &= \beta_{n-1} f_{n-1} + \alpha_n f_n.
\end{aligned} \qquad (104.9)
$$

Si en una sucesión de potencias faltan n vectores linealmente independientes, entonces, aplicando el proceso de ortogonalización, llegamos a que $f_{r+1} = 0$ para cierto $r < n$. Tomemos un vector nuevo u y formemos el vector

$$
v = u - \sum_{j=1}^{r} \eta_j f_j,
$$

determinando los coeficientes η_j de las condiciones de ortogonalidad del vector v a la izquierda respecto de los vectores $f_1, f_2, \ldots, f_r$. Construyamos una sucesión de potencias engendrada por el vector v. Es fácil probar que todo vector de esta sucesión es también ortogonal a la izquierda respecto de los vectores $f_1, f_2, \ldots, f_r$. Dicha propiedad la posee por construcción el vector v. Supongamos que la hemos demostrado para todos los vectores $v, Av, \ldots, A^k v$. En este caso, teniendo presentes las correlaciones (104.9) y la igualdad (104.7), obtenemos para el operador A

$$
\begin{aligned}
(A^{k+1}v, f_l) = (A^k v, A^* f_l) = (A^k v, (\alpha E + \beta A) f_l) = \\
= \bar{\alpha}(A^k v, f_l) + \bar{\beta}(A^k v, \beta_{l-1} f_{l-1} + \alpha_l f_l + f_{l+1}) = 0.
\end{aligned}
$$

Aplicando el proceso de ortogonalización a la nueva sucesión, construiremos un sistema de vectores linealmente independientes $q_1, q_2, \ldots, q_s$ que son seudoortogonales entre sí y a la totalidad de los vectores $f_1, f_2, \ldots, f_r$. Si $r + s < n$, continuaremos el proceso de construcción de la base hasta que se haya construido la base de todo el espacio. En este caso todo el espacio se dividirá en la suma directa

de los subespacios invariantes, y la matriz del operador A en dicha base será de forma celular diagonal con células tridiagonales del tipo (104.8).

Ejercicios.

1. Demuéstrese que el polinomio mínimo que anula el vector a es un divisor del polinomio característico.

2. Demuéstrese que el polinomio mínimo que anula el vector a es único, salvo un factor escalar.

3. Demuéstrese que si el operador A es hermitiano y el espacio K_n, unitario, entonces las fórmulas (104.8) adquieren la forma:

$$\alpha_i = \frac{(Al_i, l_i)}{(l_i, l_i)},$$

$$\beta_{i-1} = \frac{(Al_i, l_{i-1})}{(l_{i-1}, l_{i-1})} = \frac{(l_i, Al_{i-1})}{(l_{i-1}, l_{i-1})} = \frac{(l_i, l_i)}{(l_{i-1}, l_{i-1})} > 0.$$

4. Demuéstrese que en las condiciones del ejercicio 3 existe tal matriz diagonal D que para la matriz A_l de (104.8) la matriz $D^{-1}A_lD$ será real simétrica tridiagonal.

5. Demuéstrese que, cumplidas las condiciones (101.7), las matrices de las formas bilineales (Ax, y), (x, Ay) en la base $l_1, \ldots, l_n$ son derechas casi triangulares.

6. Demuéstrese que en las condiciones del ejercicio 3 las matrices de las formas bilineales (Ax, y), (x, Ay) en la base $l_1, \ldots, l_n$ son hermitianas tridiagonales.

7. Demuéstrese que si el espacio K_n es degenerado, entonces, con ayuda de los procesos (103.5), (103.8), puede construirse la base seudoortogonal de un subespacio regular de dimensión máxima.

105. Métodos de direcciones conjugadas

La construcción de los sistemas de vectores ortogonales, seudoortogonales y otros, especialmente a la base del empleo de las sucesiones de potencias, ofrece grandes posibilidades en la elaboración de toda una serie de métodos numéricos para solucionar las ecuaciones del tipo

$$Ax = b, \qquad (105.1)$$

donde A es un operador que actúa en el espacio lineal K_n, mientras que b es un vector prefijado y x, el vector buscado.

Ya nos hemos referido reiteradamente a los diferentes aspectos de este problema. Ahora daremos a conocer un gran grupo de métodos numéricos para la resolución de la ecuación (105.1) a los que se ha atribuido el nombre general de *métodos de direcciones conjugadas*. Todos ellos están basados en los procesos de ortogonalización de sucesiones de potencias. Con el fin de simplificar la exposición, supongamos que el operador A es regular y, por lo tanto, la ecuación

(105.1) tiene siempre la única solución. Convengamos en considerar que el espacio K_n es complejo y el producto escalar en él viene dado mediante una forma bilineal hermitiana simétrica y definido positiva, es decir, K_n es un espacio unitario.

Tomemos unos operadores regulares C, B cualesquiera y sea $s_1, \ldots, s_n$ un sistema de vectores CAB-seudoortogonal, es decir

$$(CABs_l, s_l) \neq 0, \qquad (CABs_l, s_k) = 0, \quad k < l,$$

para todo l. Designemos con x_0 al vector inicial y sea

$$x = x_0 + B\sum_{j=1}^{n} a_j s_j,$$

$$x_l = x_0 + B\sum_{j=1}^{l} a_j s_j, \tag{105.2}$$

$$r_l = Ax_l - b.$$

Entonces, de las correlaciones

$$x_l = x_{l-1} + a_l B s_l \tag{105.3}$$

se infiere que

$$r_l = r_{l-1} + a_l A B s_l. \tag{105.4}$$

Es fácil mostrar que para el sistema CAB-seudoortogonal $s_1, \ldots, s_n$ tienen lugar las igualdades

$$(Cr_l, s_k) = 0, \qquad 1 \leqslant k \leqslant l. \tag{105.5}$$

En efecto,

$$r_l = Ax_l - b = A(x_l - x) = -\sum_{j=l+1}^{n} a_j ABs_j$$

y, además,

$$(Cr_l, s_k) = -\sum_{j=l+1}^{n} a_j (CABs_j, s_k) = 0$$

para cualesquiera $k \leqslant l$.

Consideraremos que el sistema de vectores $s_1, \ldots, s_n$ se construye paralelamente con el sistema $r_0, \ldots, r_{n-1}$, con ayuda del proceso de su CAB-seudoortogonalización. Hagamos $s_1 = r_0$ y para todo l tendremos

$$s_{l+1} = r_l + \sum_{k=1}^{l} \beta_{k,\,l+1} s_k. \tag{105.6}$$

Las condiciones de CAB-ortogonalidad del vector s_{l+1} a la izquierda respecto de los vectores $s_1, \ldots, s_l$ proporcionan, como siempre, un sistema triangular izquierdo para determinar los coeficientes $\beta_{k,\,l+1}$. En este caso, r_k es una combinación lineal de los vectores $s_1, \ldots, s_{k+1}$. Por consiguiente, el producto escalar (Cr_l, r_k) es una combinación lineal de los números $(Cr_l, s_1), \ldots, (Cr_l, s_{k+1})$ y es igual

a cero, de conformidad con (105.5) para $k < i$, es decir,

$$(Cr_i, r_k) = 0, \qquad k < i. \tag{105.7}$$

Esto significa que si $(Cr_i, r_i) \neq 0$ para i cualquiera, entonces la sucesión de vectores r_i es C-seudoortogonal. En un espacio lineal n-dimensional un sistema C-seudoortogonal no puede contener más de n vectores no nulos. Por esta razón, en cierto paso del proceso de cálculo uno de los residuos se hará nulo y obtendremos la solución exacta de la ecuación (105.1).

Para realizar el proceso, es necesario determinar los coeficientes a_i de (105.2) y $\beta_{k,i+1}$ de (105.6). Los coeficientes a_i se hallan siempre de un modo simple. Conforme a (105.4), (105.5), (105.7), tenemos

$$a_i = -\frac{(Cr_{i-1}, r_{i-1})}{(CABs_i, r_{i-1})} = -\frac{(Cr_{i-1}, r_{i-1})}{(CABs_i, s_i)}. \tag{105.8}$$

En el caso general, los coeficientes $\beta_{k,i+1}$ se calculan de una manera mucho más compleja. Sin embargo, si los operadores A, B, C están ligados entre sí por la correlación

$$(CABC^{-1})^* = \alpha E + \beta AB \tag{105.9}$$

para ciertos números α, β, entonces entre todos los coeficientes $\beta_{k,i+1}$ sólo $\beta_{i,i+1}$ puede ser distinto de cero. Supongamos que

$$s_{i+1} = r_i + b_i s_i. \tag{105.10}$$

El coeficiente b_i se determina unívocamente de la condición de CAB-ortogonalidad del vector s_{i+1} a la izquierda respecto del vector s_i, lo que nos da, al tomar en consideración (105.9), (105.10)

$$b_i = -\frac{(CABr_i, s_i)}{(CABs_i, s_i)} = -\bar{\beta}\,\frac{(Cr_i, ABs_i)}{(CABs_i, s_i)}. \tag{105.11}$$

Supongamos que al construir la sucesión de los vectores s_i según las fórmulas (105.10), (105.11), hemos mostrado que la sucesión $s_1, \ldots, s_i$ forma un sistema CAB-seudoortogonal. Esto es justo, a ciencia cierta, cuando $i = 2$. Teniendo presente (105.9), obtenemos de (105.4)—(105.7), para $k < i$, que

$$\begin{aligned}(CABs_{i+1}, s_k) &= (CABr_i, s_k) + b_i\,(CABs_i, s_k) = \\ &= ((CABC^{-1})\,Cr_i, s_k) = (Cr_i, (CABC^{-1})^* s_k) = \\ &= (Cr_i, (\alpha E + \beta AB)s_k) = \bar{\alpha}\,(Cr_i, s_k) + \bar{\beta}\,(Cr_i, ABs_k) = \\ &= \bar{\beta}\left(Cr_i, \frac{1}{a_k}(r_k - r_{k-1})\right) = \frac{\bar{\beta}}{\bar{a}_k}\,\{(Cr_i, r_k) - (Cr_i, r_{k-1})\} = 0.\end{aligned}$$

De este modo, si se cumple la correlación (105.9), la resolución de la ecuación operacional (105.1) puede realizarse a base de la

siguiente sugestión:

$$
\begin{aligned}
e_1 &= r_0,\\
r_i &= r_{i-1} + a_i ABe_i,\\
e_{i+1} &= r_i + b_i e_i,\\
x_i &= x_{i-1} + a_i Be_i.
\end{aligned}
\qquad (105.12)
$$

Aquí, x_0 es un vector inicial arbitrario y los coeficientes a_i, b_i se calculan de acuerdo con (105.8), (105.11). Al designar $u_i = Be_i$, el proceso será:

$$
\begin{aligned}
u_1 &= Br_0,\\
r_i &= r_{i-1} + a_i Au_i,\\
u_{i+1} &= Br_i + b_i u_i,\\
x_i &= x_{i-1} + a_i u_i
\end{aligned}
$$

siendo para este caso

$$
a_i = -\frac{(Cr_{i-1}, e_i)}{(CAu_i, e_i)} = -\frac{(B^{-1*}Cr_{i-1}, u_i)}{(B^{-1*}CAu_i, u_i)},
$$

$$
b_i = -\frac{(B^{-1*}CABr_i, u_i)}{(B^{-1*}CAu_i, u_i)} = -\frac{(Cr_i, Au_i)}{(B^{-1*}CAu_i, u_i)}.
$$

Estos procesos llevan el nombre de *métodos de direcciones conjugadas*.

De las fórmulas (105.4), (105.10) concluimos que los vectores r_i, e_{i+1} son combinaciones lineales de los vectores de una misma sucesión de potencias

$$
r_0,\ ABr_0,\ \ldots,\ (AB)^i r_0. \qquad (105.13)
$$

Más aún, se han obtenido de ésta con ayuda de C- y CAB-seudoortogonalización, respectivamente. De este hecho se deducen unos corolarios de importancia exclusiva.

Si en la descomposición del vector r_0 según la base canónica de Jordan del operador AB no todos los componentes están presentes, entonces la anulación del residuo sucederá antes que llegue el n-ésimo paso. El proceso se termina con una rapidez singular, si el operador AB es de estructura simple y tiene un gran número de valores propios coincidentes. A saber, si en la descomposición del vector r_0 según los valores propios de la matriz AB las componentes no nulas corresponden a m valores propios distintos dos a dos, entonces $r_m = 0$.

De acuerdo con el teorema 104.1, para los vectores e_i, r_i deben tener lugar las correlaciones de tres términos del tipo (104.5). Pueden obtenerse éstos directamente de (105.4), (105.10). A saber,

$$
\begin{aligned}
e_{i+1} &= a_i ABe_i + (1 + b_i)e_i - b_{i-1}e_{i-1}, \quad i > 1,\\
r_{i+1} &= a_{i+1} ABr_i + \left(1 + \frac{b_i a_{i+1}}{a_i}\right) r_i - \frac{b_i a_{i+1}}{a_i} r_{i-1}, \quad i \geqslant 1.
\end{aligned}
\qquad (105.14)
$$

De aquí pueden obtenerse también otras correlaciones. Por ejemplo,

$$x_{i+1} = x_{i-1} + \omega_{i+1}(\alpha_i B r_i + x_i - x_{i-1}),$$

donde ω_{i+1}, α_i son unos números elegidos de modo adecuado.

En relación con lo dicho respecto de la expresión (105.13), fijemos nuestra atención en la siguiente peculiaridad de la condición (105.9). A primera vista, se diferencia de las condiciones del tipo (101.7). Sin embargo, al tomar en consideración (101.6), resulta fácil mostrar que la condición (105.9) es, de hecho, también una condición del tipo (101.7) y, además, simultáneamente respecto a dos productos escalares $(CABx, y)$ y (Cx, y). En efecto, observemos que el operador conjugado en (105.9) está ligado con el producto escalar principal del espacio unitario, mientras que la ortogonalidad de los vectores r_i, r_j se cargan respecto de los productos escalares $(CABx, y)$ y (Cx, y), respectivamente. Por ello

$$_{CAB}(AB)^* = (CAB\cdot AB\cdot(CAB)^{-1})^* = (CABC^{-1})^* = \alpha E + \beta AB,$$

$$_{C}(AB)^* = (CABC^{-1})^* = \alpha E + \beta AB.$$

La realización de los métodos de direcciones conjugadas puede ser obstaculizada sólo por el hecho de que uno de los productos escalares, $(CABs_i, s_i)$ o (Cr_{i-1}, r_{i-1}), se anula antes de que se reduzca a cero el residuo. Si $(CABs_i, s_i) = 0$, los coeficientes a_i, b_i no pueden calcularse. En cambio, si $(Cr_{i-1}, r_{i-1}) = 0$, esto tendrá por resultado que el coeficiente a_i será nulo, mientras que los residuos no nulos r_{i-1}, r_i coincidirán y, como consecuencia, tendrá lugar la igualdad $(CABs_{i+1}, s_{i+1}) = 0$. Se puede evitar esta situación mediante la elección del nuevo vector inicial x_0. Si los operadores CAB y C son definidos positivos, las degeneraciones citadas son imposibles y el proceso fluye sin complicaciones. Si el operador CAB es definido positivo, los métodos de direcciones conjugadas adquieren adicionalmente unas nuevas propiedades interesantes.

Sobre el grado en que el vector x_i se aproxima a la solución de la ecuación (105.1) podemos juzgar por la pequeñez del cuadrado de una norma de la diferencia $e = x - z$. Con este fin resulta cómodo utilizar las así llamadas *funcionales generalizadas de errores* del tipo (Re, e), donde R es cualquier operador definido positivo, por ejemplo, el operador $B^{-1*}CA$. El operador dado será definido positivo, puesto que está asociado al operador CAB mediante la relación $B^{-1*}CA = B^{-1*}(CAB)B^{-1}$. Es válido el

TEOREMA 105.1. *Si el operador CAB es definido positivo, entre todos los vectores del tipo $z = x_0 + Bv$, donde v pertenece a la cápsula lineal de los vectores $r_1, \ldots, r_i$, el vector x_i da el mínimo de la funcional generalizada de errores*

$$\varphi(z) = (B^{-1*}CAe, e)$$

DEMOSTRACIÓN. Puesto que el operador CAB es definido positivo, el sistema de los vectores s_i será CAB-ortogonal. Representemos el vector z en forma de una descomposición, por analogía con la descomposición (105.2) para el vector x:

$$z = x_0 + B \sum_{j=1}^{i} h_j s_j.$$

Tenemos

$$\varphi(z) = (B^{-1*}CA(x-z),\ x-z) =$$

$$= \Big(B^{-1*}CAB\Big(\sum_{j=1}^{m} \alpha_j s_j - \sum_{j=1}^{i} h_j s_j\Big),\ B\Big(\sum_{j=1}^{m} \alpha_j s_j - \sum_{j=1}^{i} h_j s_j\Big)\Big) =$$

$$= \Big(CAB\Big(\sum_{j=1}^{m} \alpha_j s_j - \sum_{j=1}^{i} h_j s_j\Big),\ \sum_{j=1}^{m} \alpha_j s_j - \sum_{j=1}^{i} h_j s_j\Big) =$$

$$= \sum_{j=1}^{i} |\alpha_j - h_j|^2 (CABs_j, s_j) + \sum_{j=i+1}^{m} |\alpha_j|^2 (CABs_j, s_j).$$

De aquí concluimos que el mínimo de la funcional de errores se alcanza para $h_j = \alpha_j$, $j \leq i$, es decir, para $z = x_i$.

La funcional de errores no puede determinarse en los cálculos prácticos, puesto que depende de la solución x que se desconoce. No obstante, dicha funcional sólo se diferencia de otra funcional en un sumando constante:

$$\psi(z) = (B^{-1*}CAz,\ z) - 2\operatorname{Re}(B^{-1*}Cb,\ z)$$

que ya puede calcularse. En efecto,

$$\varphi(z) = (B^{-1*}CA(x-z),\ x-z) =$$
$$= (B^{-1*}CAx,\ x) - (B^{-1*}CAx,\ z) - (B^{-1*}CAz,\ x) +$$
$$+ (B^{-1*}CAz,\ z) = (B^{-1*}CAz,\ z) - (B^{-1*}Cb,\ z) -$$
$$- \overline{(B^{-1*}Cb,\ z)} + (B^{-1*}Cb,\ x) = \psi(z) + (B^{-1*}Cb,\ x).$$

Demos a conocer, en fin, algunas clases de los operadores, para los cuales se cumplen las condiciones (105.9).

1. Todos los operadores A, B, C son hermitianos, siendo $B = C$. La condición (105.9) se cumple para $\alpha = 0$, $\beta = 1$:

$$(CABC^{-1})^* = (BA)^* = A^*B^* = 0 \cdot E + 1 \cdot AB.$$

2. Los operadores CAB y C son hermitianos. La condición (105.9) se cumple para $\alpha = 0$, $\beta = 1$:

$$(CABC^{-1})^* = C^{-1*}(CAB)^* = C^{-1}CAB = 0 \cdot E + 1 \cdot AB.$$

3. El operador C se conmuta con AB; el operador AB es normal y su espectro se halla en una línea recta. La última condición signi-

lica que $AB = \gamma E + \delta H$ para cierto operador hermitiano H. Ahora hallemos

$$(CABC^{-1})^* = (CC^{-1}AB)^* = (AB)^* = (\gamma E + \delta H)^* =$$

$$= \bar{\gamma}E + \bar{\delta}H = \frac{\bar{\gamma}\delta - \bar{\delta}\gamma}{\delta} E + \frac{\bar{\delta}}{\delta}(\gamma E + \delta H) = + \frac{2i \operatorname{Im}(\bar{\gamma}\delta)}{\delta} E + \frac{\bar{\delta}}{\delta} AB.$$

4. Representemos el operador A en la forma $A = M + N$, donde $M = M^*$, $N = -N^*$. Si el operador M es regular, hacemos $B = C = M^{-1}$. La condición (105.9) se cumple cuando $\alpha = 2$, $\beta = -1$:

$$(CABC^{-1})^* = (M^{-1}(M + N))^* = (M - N)M^{-1} =$$
$$= 2E - (M + N)M^{-1} = 2 \cdot E - 1 \cdot AB.$$

5. Si el operador N en la descomposición $A = M + N$ es regular, hacemos $B = C = N^{-1}$. La condición (105.9) se cumple para $\alpha = 2$, $\beta = -1$:

$$(CABC^{-1})^* = (N^{-1}(M + N))^* = -(M - N)N^{-1} =$$
$$= 2E - (M + N)N^{-1} = 2 \cdot E - 1 \cdot AB.$$

Ejercicios.

1. Demuéstrese que la matriz de la forma bilineal $(CABx, y)$ es:
triangular derecha en la base $s_1, \dots, s_n$;
casi triangular derecha en la base $r_1, \dots, r_{n}$;
triangular izquierda en las bases $s_1, \dots, s_n$ y $r_1, \dots, r_{n}$, si el operador C es hermitiano.

2. ¿Cómo varía la forma de una matriz de la forma bilineal $(CABx, y)$ en el ejercicio 1, si el operador CAB es hermitiano?

3. Demuéstrese que la matriz de la forma bilineal (Cx, y) es:
triangular derecha en la base $r_0, \dots, r_{n-1}$;
triangular derecha en las bases $r_0, \dots, r_{n-1}$ y $s_1, \dots, s_n$;
triangular derecha en las bases $ABs_1, \dots, ABs_n$ y $s_1, \dots, s_n$;
casi triangular derecha en las bases $ABr_0, \dots, ABr_{n-1}$ y $r_0, \dots, r_{n-1}$.

4. ¿Cómo varía la forma de una matriz de la forma bilineal (Cx, y) en los ejercicios 3, si el operador C es hermitiano?

5. Demuéstrese que si la condición (105.9) se sustituye por otra:

$$(CABC^{-1})^* = \alpha_0 E + \alpha_1 AB + \dots + \alpha_p (AB)^p, \qquad p \geqslant 1,$$

la correlación (105.10) tendrá por expresión

$$s_{i+1} = r_i + b_{ii}s_i + b_{i,i-1}s_{i-1} + \dots + b_{i,i-p+1}s_{i-p+1}.$$

6. Demuéstrese que

$$a_i = -\frac{(Cr_{i-1}, s_{i-1})}{(CABs_i, s_i)}.$$

7. Demuéstrese que si los operadores CAB y C son hermitianos, se verifica

$$b_i = \frac{(Cr_i, r_i)}{(Cr_{i-1}, r_{i-1})}.$$

8. Demuéstrese que si los operadores CAB y C son hermitianos y, además, definidos positivos, entonces $a_i < 0$, $b_i > 0$ para todo valor de i.

9. Demuéstrese que la matriz del operador AB tiene en una base de los vectores $s_1, \ldots, s_n$ ó $r_0, \ldots, r_{n-1}$ la forma tridiagonal.
10. ¿Cómo se realizan los métodos de direcciones conjugadas en el caso cuando el operador A es degenerado?

§ 106. Variantes principales

Consideraremos las variantes más conocidas de los métodos de direcciones conjugadas. En el aspecto teórico todas ellas se encajan dentro del esquema descrito anteriormente (105.12). Sin embargo, los cálculos prácticos se efectúan, a veces, a base de los algoritmos algo diferentes de éste.

Método de gradientes conjugados. En este caso el operador A es hermitiano y, además, definido positivo; $B = C = E$; la condición (105.9) se cumple para $\alpha = 0$, $\beta = 1$. Debido a que los operadores $CAB = A$ y $C = E$ son definidos positivos se garantiza la ausencia de degeneraciones en el proceso de cálculo. En cada etapa del método se minimiza la funcional de errores provista de la matriz A. El esquema de cálculo de este método es como sigue

$$s_1 = r_0,$$
$$r_i = r_{i-1} + a_i A s_i,$$
$$s_{i+1} = r_i + b_i s_i,$$
$$x_i = x_{i-1} + a_i s_i,$$

donde

$$a_i = -\frac{(r_{i-1}, r_{i-1})}{(As_i, r_{i-1})} = -\frac{(r_{i-1}, s_i)}{(As_i, s_i)} < 0,$$
$$b_i = -\frac{(r_i, As_i)}{(As_i, s_i)} = \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} > 0.$$

Al aplicar el método de gradientes conjugados los vectores r_i forman un sistema ortogonal y los vectores s_i, un sistema A-ortogonal.

Método de iteraciones AA^*-minimales. En este caso el operador A es regular arbitrario; $B = A^*$, $C = E$; la condición (105.9) se cumple para $\alpha = 0$, $\beta = 1$. El hecho de que los operadores $CAB = AA^*$ y $C = E$ son definidos positivos garantiza la ausencia de las degeneraciones en el proceso de cálculo. En cada etapa del método se minimiza la funcional de errores con la matriz E, es decir, el cuadrado de la norma euclidiana del mismo error. El esquema de cálculo es como sigue

$$u_1 = A^* r_0,$$
$$r_i = r_{i-1} + a_i A u_i,$$
$$u_{i+1} = A^* r_i + b_i u_i,$$
$$x_i = x_{i-1} + a_i u_i,$$

donde

$$a_i = -\frac{(r_{i-1}, r_{i-1})}{(Au_i, r_{i-1})} = -\frac{(r_{i-1}, r_{i-1})}{(u_i, u_i)} < 0,$$

$$b_i = -\frac{(r_i, Au_i)}{(u_i, u_i)} = \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} < 0.$$

Al aplicar el método de iteraciones AA^*-minimales los vectores r_i y u_i forman sistemas ortogonales.

Método de iteraciones A^*A-minimales. En este caso el operador A es regular arbitrario; $B = A^*$, $C = AA^*$; la condición (105.9) se cumple para $\alpha = 0$, $\beta = 1$. El hecho de que los operadores $CAB = (AA^*)^*$ y $C = AA^*$ son definidos positivos garantiza la ausencia de las degeneraciones en el proceso de cálculo. En cada etapa del método se minimiza la funcional de errores prevista de la matriz A^*A, es decir, el cuadrado de la norma euclidiana del vector del residuo. El esquema de cálculo es como sigue

$$u_1 = A^*r_0,$$

$$r_i = r_{i-1} + a_iAu_i,$$

$$u_{i+1} = A^*r_i + b_iu_i,$$

$$x_i = x_{i-1} + a_iu_i,$$

de donde

$$a_i = -\frac{(A^*r_{i-1}, A^*r_{i-1})}{(Au_i, Au_i)} < 0, \qquad b_i = \frac{(A^*r_i, A^*r_i)}{(A^*r_{i-1}, A^*r_{i-1})} > 0.$$

Al aplicar el método de iteraciones A^*A-minimales los vectores A^*r_i y Au_i forman sistemas ortogonales.

Método de descomposición hermitiana completa. En este caso el operador A es regular arbitrario. Representémoslo en forma de una suma $A = M + N$, donde $M = M^*$, $N = -N^*$. En el caso en que M o N sea regular, ponemos $B = C = M^{-1}$ o $B = C = N^{-1}$, respectivamente. La condición (105.9) se cumple para $\alpha = 2$, $\beta = -1$. Si el operador M (o iN) es de signo constante, el proceso se realiza sin degeneraciones. Sea, por ejemplo, $M > 0$ y $B = C = M^{-1}$. El operador C será definido positivo, por lo tanto $(Cz, z) = (M^{-1}z, z) > 0$, para cualquier vector z no nulo. Consideremos ahora el operador $CAB = M^{-1} + M^{-1}NM^{-1}$. Para cualquier $z \neq 0$ tenemos

$$(CABz, z) = (M^{-1}z, z) + (M^{-1}NM^{-1}z, z) \neq 0,$$

puesto que el primer producto escalar del segundo miembro de la igualdad es real y positivo, por ser definido positivo el operador, mientras que el segundo producto escalar es imaginario puro, en virtud de que el operador $M^{-1}NM^{-1}$ es antihermitiano. Para el caso en que $B = C = M^{-1}$, el esquema de cálculo del método es

como sigue

$$Mu_1 = r_0,$$
$$r_i = r_{i-1} + a_i A u_i,$$
$$Mv_i = r_i,$$
$$u_{i+1} = v_i + b_i u_i,$$
$$x_i = x_{i-1} + a_i u_i,$$

donde

$$a_i = -\frac{(r_{i-1}, u_i)}{(Au_i, u_i)}, \qquad b_i = \frac{(v_i, Au_i)}{(Au_i, u_i)}.$$

Si $B = C = N^{-1}$, el esquema de cálculo y las fórmulas para los coeficientes a_i, b_i quedan en vigor, a excepción de que M se sustituye por N.

Método de descomposición hermitiana incompleta. En este caso el operador A es hermitiano, definido positivo. Representémoslo en forma de una suma $A = M + N$, donde $M = M^*$, $N = N^*$. Si M es regular, ponemos $B = C = M^{-1}$. La condición (105.9) se cumple cuando $\alpha = 0$, $\beta = 1$. Si M es un operador definido positivo, el proceso se realiza sin degeneraciones. En cada etapa del método se minimiza la funcional de errores provista de la matriz A. El esquema de cálculo queda el mismo que en el caso antecedente.

Aceleración del proceso de cálculo. Como ya se ha notado anteriormente, los métodos de direcciones conjugadas permiten hallar la solución con una rapidez singular, si el operador AB cuenta con pocos valores propios distintos dos a dos. Esta singularidad sirve de base para diferentes procedimientos que tienen por objeto acelerar la resolución de la ecuación (105.1), empleando la siguiente idea.

Supongamos que el operador A puede representarse en forma de una suma $A = M + N$, donde el operador M determina la parte «principal» del operador A y admite, además, una resolución sencilla de las ecuaciones del tipo (105.1) con el operador M en el primer miembro. Ahora, en lugar de la ecuación (105.1) resolveremos la ecuación

$$(E + NM^{-1})\, y = b, \tag{106.1}$$

donde $Mx = y$. Si, en algún sentido razonable, el operador M es próximo al operador A, entonces la mayoría de los valores propios del operador N y, por tanto, del operador NM^{-1} serán próximos a cero o iguales a cero. La aplicación de los métodos de direcciones conjugadas a la ecuación (106.1) conduce, en el caso dado, a su rápida resolución.

Ha de ser notado que precisamente esta idea es la base para crear un método de descomposición hermitiana incompleta, el cual en muchos casos resulta más eficaz que el de gradientes en la variante

clásico. Todo depende de cuán exitosa ha sido la descomposición del operador A.

No es nuestra intención detenernos detalladamente en los esquemas de cálculo para los procesos de aceleración, pues dependen demasiado del empleo de unas u otras peculiaridades del operador A.

Ejercicios.

1. ¿En qué condiciones resulta conveniente aplicar una u otra variante del método de direcciones conjugadas?

2. ¿Qué número de iteraciones se deben cumplir, realizando diversas variantes de los métodos de direcciones conjugadas para un operador A del tipo $E + R$, donde R es de rango r?

3. Considerando el operador A una matriz, calcúlese el número de operaciones aritméticas que se deben cumplir al resolver sistemas de las ecuaciones algebraicas lineales por los métodos de direcciones conjugadas.

4. La matriz del operador A es hermitiana y se diferencia de la matriz triangular por un número pequeño de elementos. ¿Cuál de las variantes de los métodos de direcciones conjugadas es la más conveniente para el empleo en el caso dado?

5. Sea $P_1(t)$, $P_2(t)$, . . . una sucesión de polinomios. Elijamos un vector x_0 y construyamos una sucesión de vectores x_0, x_1, . . ., rigiéndonos por la regla

$$x_{k+1} = x_k - BP_k(AB)(Ax_k - b), \qquad k \geq 0. \tag{106.2}$$

¿Cómo varían las descomposiciones de los residuos r_0, r_1, . . . según la base canónica de Jordan del operador AB, cuando k crece en función de la elección de la sucesión de polinomios?

6. ¿Cómo debe utilizarse la sucesión (106.2) con el fin de construir un vector inicial para los métodos de direcciones conjugadas el cual asegure la obtención de la solución por una cantidad menor de iteraciones?

7. ¿Cuáles de los sistemas de vectores en cada una de las variantes concretas de los métodos de direcciones conjugadas son, salvo una normalización, A-seudoduales?

§ 107. Ecuaciones operacionales y seudodualidad

Los métodos de direcciones conjugadas no son únicos entre aquellos que se emplean para la resolución de la ecuación operacional

$$Ax = b, \tag{107.1}$$

y están basados en la aplicación de las formas bilineales. Enormes posibilidades para crear estos métodos ofrece la construcción de los sistemas de vectores, duales o seudoduales respecto a cierta forma bilineal, vinculada con el operador A de la ecuación (107.1).

Volvamos a considerar que el operador A es regular y actúa en un espacio unitario K_n. Examinemos una forma bilineal (Ax, y) y supongamos que para ésta se han obtenido, de una u otra manera, los

sistemas de vectores $u_1, u_2, \ldots, u_n$ y $v_1, v_2, \ldots, v_n$ que son A-seudoduales, salvo una normalización, es decir,

$$(Au_i, v_i) \neq 0, \qquad (Au_i, v_k) = 0, \quad k < i, \tag{107.2}$$

para todo valor de $i \leqslant n$. Probemos que el conocimiento de los sistemas A-seudoduales de vectores permite construir un proceso de búsqueda de la solución de la ecuación (107.1).

Elijamos un vector arbitrario x_0. Como los sistemas A-seudoduales son linealmente independientes, existe la descomposición

$$x = x_0 + \sum_{j=1}^{n} \alpha_j u_j. \tag{107.3}$$

Si

$$x_i = x_0 + \sum_{j=1}^{i} \alpha_j u_j,$$

entonces, por analogía con (105.3), (105.4), tenemos

$$x_i = x_{i-1} + \alpha_i u_i, \qquad r_i = r_{i-1} + \alpha_i A u_i. \tag{107.4}$$

Luego,

$$r_i = Ax_i - b = A(x_i - x) = -\sum_{j=i+1}^{n} \alpha_j A u_j,$$

y, en plena concordancia con las segundas condiciones (107.2), encontramos que

$$(r_i, v_k) = -\sum_{j=i+1}^{n} \alpha_j (Au_j, v_k) = 0$$

para todo valor de $k \leqslant i$. Así pues,

$$(r_i, v_k) = 0 \tag{107.5}$$

para $k \leqslant i$. Esto nos permite determinar los coeficientes α_i de (107.4), a saber

$$\alpha_i = \frac{(r_{i-1}, v_i)}{(Au_i, v_i)}. \tag{107.6}$$

De acuerdo con la primera condición (107.2), el denominador en el segundo miembro de (107.6) es distinto de cero.

De (107.5) se deduce que el vector r_n será ortogonal a la izquierda y, por ser simétrico el producto escalar, simplemente ortogonal a los vectores linealmente independientes $v_1, v_2, \ldots, v_n$, es decir, $r_n = 0$, mientras que x_n es la solución de la ecuación (107.1).

Los métodos descritos de resolución de la ecuación (107.1) llevan el nombre general de métodos de direcciones duales. El número de diferentes métodos es infinito en pleno sentido de esta palabra, puesto que existe un número infinito de distintos pares A-seudoduales de los sistemas de vectores. Los métodos de direcciones conjugadas, considerados anteriormente, forman parte de este grupo.

En el caso general, para los métodos de direcciones duales no existe ningún análogo del teorema 105.1, incluso cuando el operador A sea definido positivo. La ley de variación de los errores $e_k = x - x_k$ en estos métodos describe el teorema siguiente que, aunque débil, es sin embargo útil.

TEOREMA 107.1 *Sea P_k un operador de proyección sobre el subespacio tendido sobre los vectores $u_1, \ldots, u_k$ paralelamente a un subespacio tendido sobre los vectores $u_{k+1}, \ldots, u_n$. Entonces*

$$e_k = (E - P_k)\, e_0. \tag{107.7}$$

DEMOSTRACIÓN. De acuerdo con la fórmula (107.3), tenemos la siguiente descomposición del vector inicial x_0 para el error e_0:

$$e_0 = x - x_0 = \sum_{j=1}^{n} \alpha_j u_j.$$

Pero, por definición del operador de proyección,

$$P_k e_0 = \sum_{j=1}^{k} \alpha_j u_j.$$

El segundo miembro de esta igualdad no es otra cosa que $x_k - x_0$. Por esto

$$P_k e_0 = x_k - x_0 = (x - x_0) - (x - x_k) = e_0 - e_k,$$

lo que demuestra la afirmación del teorema.

Unos resultados interesantes relacionados con los sistemas A-seudoduales pueden obtenerse, considerando la interpretación matricial de los métodos descritos.

Consideraremos que el espacio K_n no es sólo unitario, sino también aritmético, que es admisible en virtud de que los espacios lineales de dimensión finita son isomorfos. Todos los razonamientos realizados quedan en vigor, sólo cambia la terminología: la ecuación (101.7) pasa a ser un sistema de ecuaciones algebraicas lineales, los operadores se sustituyen por matrices y por vectores se entienden los vectores-columna. Designaremos mediante U (V) una matriz cuyas columnas son los vectores $u_1, \ldots, u_n$ ($v_1, \ldots, v_n$). En tales circunstancias el hecho de que estos vectores satisfacen las correlaciones (107.2) significa que la matriz

$$C = V^* A U$$

es regular triangular izquierda. De aquí se desprende la siguiente descomposición de la matriz A en factores:

$$A = V^{-1*} C U^{-1}. \tag{107.8}$$

Así pues, el hecho de que se conocen los sistemas de vectores, A-seudoduales, salvo una normalización, permite resolver el sistema de ecuaciones algebraicas lineales (101.7) con la evaluación de errores (107.7) y, por otra parte, obtener la descomposición (107.8) de la

matriz A en factores, entre los cuales hay uno triangular. Demostremos que la afirmación recíproca es también cierta. A saber, cualquier método de resolución de los sistemas de ecuaciones algebraicas lineales, basado en una descomposición de la matriz en factores, entre los cuales hay aunque sea un solo factor triangular, determina ciertos sistemas de vectores, A-seudoduales, salvo una normalización. Por consiguiente, al realizar estos métodos según los esquemas (107.3)—(107.6), se pueden utilizar las evaluaciones (107.7).

Examinemos la matriz P que se obtiene de una matriz unidad intercambiando sus columnas o, lo que es igual, intercambiando las filas en el orden inverso. Es fácil comprobar que la multiplicación a la derecha de la matriz C por P cambia de lugar en el orden inverso las columnas de la matriz C; la multiplicación a la izquierda de la matriz CP por P conmuta en el orden inverso las filas de la matriz CP. Por esta razón, los elementos f_{ij} de la matriz $F = PCP$ están vinculados con los elementos c_{ij} de la matriz C mediante una correlación $f_{ij} = c_{n-i+1,\, n-j+1}$.

De aquí puede sacarse toda una serie de corolarios útiles. Numeremos las diagonales de la matriz, paralelas a la principal, de abajo arriba, una tras otra, con los números $-(n-1)$, $-(n-2)$, ..., ..., 0, ..., $(n-2)$, $(n-1)$. La diagonal con el número 0 está principal. Aceptada tal numeración, los elementos de la k-ésima diagonal se determinan por la correlación $j - i = k$. Si la matriz C satisface las condiciones $c_{ij} = 0$, $k < j - i$, $j - i < l$ para ciertos números $l \leqslant k$, entonces para la matriz $F = PCP$ se cumplirán las igualdades $f_{ij} = 0$, $-l < j - i$, $j - i < -k$. Por consiguiente, en la transformación de $F = PCP$ la matriz diagonal sigue siendo diagonal, la matriz triangular derecha (izquierda) se convierte en triangular izquierda (derecha) y la matriz bidiagonal derecha (izquierda), en bidiagonal izquierda (derecha), etc.

Supongamos ahora que se aplica cierto método de resolución del sistema de ecuaciones algebraicas lineales (107.1), basado en la descomposición previa de la matriz A en factores:

$$A = QCR, \tag{107.9}$$

donde la matriz C es triangular. Sin restringir esencialmente la generalidad, podemos considerar que C es triangular izquierda puesto que de lo contrario, en lugar de la descomposición (107.9) consideraríamos la descomposición

$$A = (QP)\,(PCP)\,(PR),$$

donde la matriz PCP debe ser triangular izquierda, de acuerdo con lo dicho más arriba. Las matrices buscadas U, V, que determinan los sistemas de vectores, A-seudoduales, salvo la normalización, $u_1, \ldots, u_n$ y $v_1, \ldots, v_n$, pueden definirse por las igualdades

$$U = R^{-1}, \qquad V = Q^{-1*},$$

Hemos de notar que en la descomposición (107.9), engendrada por un método numérico, las matrices Q y R, son como regla, suficientemente sencillas. Estas son, con frecuencia, matrices unitarias o rectangulares y también las matrices que difieren de las triangulares en que las filas y columnas están permutadas. Por ello, las matrices R^{-1} y Q^{-1*} se hallan sin dificultad. En todo caso, los esfuerzos totales de cálculo para su determinación son considerablemente inferiores a aquellos que son necesarios para obtener la descomposición (107.9). Estas propiedades las poseen tales métodos ampliamente conocidos como el método de Gauss, de raíces cuadradas, de Jordan, de ortogonalización, de relajaciones, de giro, etc.; los métodos basados en la reducción del sistema a la forma bidiagonal con obtención de las descomposiciones normalizadas; los métodos de direcciones conjugadas, etc.

De este modo, la mayoría de los métodos numéricos que se emplean para la resolución de las ecuaciones operacionales (107.1) en un espacio de dimensión finita son, en realidad, métodos de construcción de los sistemas A-conjugados duales de vectores. Pese a la diversidad de las formas concretas de los propios métodos, todos ellos pueden investigarse a partir de las posiciones generales, basándose siempre en el teorema 107.1.

Ejercicios.

1. Convengamos en considerar que un operador es su matriz y los vectores de un espacio, vectores columna. Demuéstrese que, aplicando el método de direcciones duales, los errores consecutivos están ligados entre sí por la correlación

$$e_k = (E - S_k)\, e_{k-1},$$

donde

$$S_k = \frac{u_k v_k^* A}{v_k^* A u_k}. \qquad (107.10)$$

2. Demuéstrese que los operadores S_k satisfacen las igualdades

$$S_k^2 = S_k, \qquad S_i S_k = 0,$$

$$S_i (E - S_k)(E - S_{k-1}) \ldots (E - S_1) = 0, \qquad i \leq k.$$

3. Demuéstrese que los operadores S_k de (107.10) y el operador P_k de (107.7) están ligados entre sí por la correlación

$$P_k = (E - S_k)(E - S_{k-1}) \ldots (E - S_1).$$

4. ¿Qué significan los operadores S_k y P_k para los métodos concretos, determinados por la descomposición (107.9)?

5. ¿Cómo varían los errores e_k para los métodos concretos, determinados por la descomposición (107.9)?

6. ¿Cuál de los métodos conocidos de resolución de los sistemas de ecuaciones algebraicas lineales no está basado en la descomposición (107.9)?

CONCLUSIÓN

En el presente manual el lector se le expone un material, suficientemente amplio, indispensable para comprender tanto la base teórica del álgebra lineal, como los métodos numéricos de ésta. No obstante, debido a las peculiaridades de diferentes programas de estudio y a la escasez eventual del tiempo que se designa a las conferencias referentes a este curso, algunos apartados de este libro pueden escaparse de la atención del lector. Es por esta razón que damos a conocer aquí la característica general de todo el curso en consideración.

El álgebra lineal, como ciencia, estudia los conjuntos de estructura especial, así como las funciones que actúan en dichos conjuntos. En general, los problemas análogos se tratan también en otras ramas de las matemáticas, por ejemplo, en el análisis matemático. La característica singular del álgebra lineal radica precisamente en que los conjuntos son siempre espacios lineales de dimensión finita y las funciones intervienen como operadores lineales.

A la consideración de las propiedades generales de los espacios lineales están dedicados los §§ 10, 13—21, las propiedades generales de los operadores lineales se examinan en los §§ 56—61, 63—74. La información, que se da en estos párrafos, puede obtenerse por varios métodos, incluso directamente sin recurrir a los conceptos y medios de investigación especiales, excepto los más elementales. Pero, uno de los conceptos adicionales merece una atención especial. Nos referimos al determinante.

Como función numérica, definida sobre los sistemas de vectores, el determinante es un ente relativamente sencillo. Sin embargo, posee varias propiedades de importancia. Estas propiedades lo convirtieron en un instrumento de amplio uso que facilita considerablemente la realización de diferentes investigaciones. Además, el determinante se emplea frecuentemente en la construcción de los métodos numéricos. Todo esto nos obligó a prestar mucha atención al concepto de determinante, al considerar en los §§ 34—42, 52 sus propiedades geométricas y algebraicas. Como instrumento de investigación el determinante se utiliza en el libro para demostrar las más diversas afirmaciones.

Otra función numérica de dos argumentos vectoriales —que es producto escalar— determina las dos clases más importantes de los espacios lineales, llamados euclídeos y unitarios. La nueva noción fundamental en dichos espacios es la de ortogonalidad. En los §§ 27—33 se estudian las propiedades de los espacios lineales, adicionales respecto del producto escalar, en los §§ 75—81, las propiedades de los operadores lineales, adicionales respecto del producto escalar.

Los sistemas de ecuaciones algebraicas lineales son de importancia exclusiva en toda la matemática, no sólo en el álgebra lineal. Al estudio de diferentes aspectos de los sistemas mencionados están dedicados los §§ 22, 45, 46, 48.

Como regla, sólo el material citado constituye la base del curso del álgebra lineal al cual se añade, como un curso independiente, el de la geometría analítica. En el presente manual la información indispensable propia al curso de la geometría analítica se da no de modo aislado, sino en conjunto con la infor-

mación correspondiente del álgebra lineal. La exposición semejante del material nos permitió lograr ciertas ventajas, a saber, se redujeron muchas demostraciones de un mismo tipo en ambos cursos, se consiguió subrayar la interpretación geométrica de tales conceptos algebraicos abstractos como el espacio lineal, el plano en un espacio lineal, el determinante, sistemas de ecuaciones algebraicas lineales, etc.

El álgebra lineal se enriquece en grado considerable con hechos nuevos, si en los espacios lineales se introducen las nociones sobre la distancia entre los vectores y límite de una sucesión de vectores. La necesidad de tal introducción está dictada también por las exigencias de los métodos numéricos. Las propiedades métricas de los espacios lineales se estudian en los §§ 49—54, las de los operadores lineales, en los §§ 82—84. Desde luego, todos estos problemas se dan comúnmente en el análisis funcional, pero, no se subrayan, como regla, muchos resultados, muy importantes para los espacios lineales de dimensión finita.

La resolución numérica de los problemas del álgebra lineal se acompaña casi siempre de la aparición de los errores de redondeo. Por esto el personal encargado de los cálculos debe darse cuenta de los cambios en las propiedades de diferentes entes del álgebra lineal a que llevan las pequeñas variaciones de los vectores y operadores. A la influencia de las pequeñas perturbaciones están dedicados los §§ 33, 87, 69.

Las propiedades de muchos entes del álgebra lineal pueden cambiarse por otras contrarias incluso cuando las perturbaciones sean pequeñas. Así por ejemplo, un sistema linealmente dependiente de vectores puede convertirse en un sistema independiente o aumentar su rango, un operador de estructura de Jordan puede pasar a ser un operador de estructura simple y un sistema compatible de ecuaciones algebraicas lineales, un sistema incompatible, etc. Todos estos hechos originan dificultades muy grandes en la resolución práctica de los problemas.

Nuestra insistente sugerencia es que el lector lea una vez más con mucha atención el § 22 y analice el ejemplo que allí se aduce. El lector debe también reflexionar profundamente sobre las preguntas al final del párrafo.

A pesar de la inestabilidad de muchas nociones del álgebra lineal, sus problemas pueden resolverse de modo bien estable. Con el fin de demostrar tal posibilidad, se ha incluido en el libro la descripción de un método estable de resolución de los sistemas de ecuaciones algebraicas lineales. El fundamento teórico y el esquema general del método citado se dan en los §§ 85, 86, 88.

La parte final del libro viene dedicada a la descripción e investigación de diferentes problemas referentes a las formas bilineales y cuadráticas. Estas funciones numéricas son de mucha importancia en el álgebra lineal y están estrechamente ligadas con la construcción de los métodos numéricos. En los §§ 90—94 se consideran las propiedades generales de las formas bilineales y la relación de sus transformaciones con las descomposiciones matriciales; en los §§ 98—101, la ampliación de la noción de ortogonalidad; en los §§ 103—107 se muestra cómo se deben emplear las formas bilineales en los procesos de cálculo.

INDICE ALFABETICO